Approximation, Proof Systems, and Correlations in a Quantum World

by

Sevag Gharibian

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Computer Science

Waterloo, Ontario, Canada, 2012

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis studies three topics in quantum computation and information: (1) The approximability of "inherently quantum" problems, (2) quantum proof systems, and (3) non-classical correlations in quantum systems. Our results in each area are summarized as follows.

Our first area of study concerns the approximability of computational problems which are complete for quantum complexity classes. In the classical setting, the study of approximation algorithms and hardness of approximation is one of the main research areas of theoretical computer science. Yet, little is known regarding approximability in the setting of quantum computational complexity. Our first result (joint work with Julia Kempe) is a polynomial-time approximation algorithm for dense instances of the canonical QMA-complete quantum constraint satisfaction problem, the local Hamiltonian problem. Our second result (joint work with Julia Kempe) goes in the opposite direction by first introducing a quantum generalization of the polynomial-time hierarchy. We then introduce problems which are not only complete for the second level of this hierarchy, but are in fact hard to approximate.

Our second area of study concerns quantum proof systems. Here, an interesting question which remains open despite much effort is whether a proof system with multiple unentangled quantum provers is equal in expressive power to a proof system with a single quantum prover (i.e. is QMA(poly) equal to QMA?). Our results here (joint work with Jamie Sikora and Sarvagya Upadhyay) study variants of this question. We first show that if each unentangled prover has logarithmic size proofs, then this is equivalent to having a single quantum prover which sends a classical proof. We then show that a variant of the class BellQMA(poly) collapses to QMA. Finally, we give an alternate proof of the fact [Harrow and Montanaro, FOCS, p. 633–642 (2010)] that the class SepQMA(m) (which is equivalent to QMA(m)) admits perfect parallel repetition. Our alternate proof is novel in that it is based on cone programming duality.

Our final area of study concerns non-classical correlations in quantum systems. Specifically, in recent years it has come to light that there appear to be genuinely quantum correlations in mixed quantum states beyond entanglement which may nevertheless prove useful from a computing and information theoretic perspective. Our first result in this area (joint work with Animesh Datta) motivates the study of such correlations by exploring possible connections to the quantum task of locking of classical correlations [DiVincenzo *et al.*, PRL 92, 067902 (2004)] and the DQC1 model of mixed-state quantum computing [Knill and Laflamme, PRL 81, 5672 (1998)]. Our second result in this area introduces a novel

scheme for quantifying non-classical correlations based on the use of local unitary operations. Our third result (joint work with Marco Piani, Gerardo Adesso, John Calsamiglia, Paweł Horodecki, and Andreas Winter) introduces and studies a protocol through which non-classical correlations in a starting system can be "activated" into distillable entanglement with an ancilla system. Surprisingly, we find that, according to the non-classicality measures derived from our protocol, mixed entangled states can be "more non-classical" than pure entangled states. Finally, our last result (joint work with Marco Piani, Gerardo Adesso, John Calsamiglia and Paweł Horodecki) continues the study of the activation protocol above by determining when the entanglement generated with the ancilla can be mapped back onto the starting state via entanglement swapping.

Acknowledgements

I'd like to congratulate myself, and thank myself, and give myself a big pat on the back.
— Dee Dee Ramone, Rock and Roll Hall of Fame induction ceremony, 2002 [4].

There are many greats in this world who have the ability to inspire and support us, whether they be artists, academics, or those we hold dear. I am indebted to the following people who have played such a role during the course of my graduate studies, without whom this thesis would not have been possible.

First, I would like to thank the readers of my thesis: Richard Cleve, Debbie Leung, Ashwin Nayak, Barbara Terhal, and John Watrous. Thank you for agreeing to take on this task; I hope it does not prove too painful.

I would like to thank my thesis advisory committee, Richard Cleve, Ashwin Nayak, and John Watrous, for their guidance and feedback, particularly in times when I have been wrong, and stubbornly so at that. I have always appreciated their constructive comments, and contrary to popular belief, feel that the more embarassing the mistake revealed by their criticism, the less likely I am to repeat the blunder in the future.

I am indebted to my supervisor, Richard Cleve, for his unfailing support over the years, whether in terms of research or at a personal level. His demand for research excellence, precision, and moral steadfastness has greatly inspired and helped guide me over the years. I may (hopefully) be leaving Waterloo having gained a Ph.D., but I will be missing a good friend.

I am also ever grateful to Julia Kempe, who has in many ways acted as a second unofficial advisor for me. Her unwavering belief in me and constant push for success has had a profound effect on my development. Coupled with her sincere hospitality, I could not imagine asking for a better host for a student on exchange. In this vein, I must also thank Oded Regev, who has also played the great host and conversation partner; his input into research projects and conference talk preparations has proven invaluable.

Though neither official nor unofficial supervisors of mine, I am also indebted to Marco Piani and John Watrous. I cannot recall any instance in which either of them has turned down an opportunity to answer one of my many questions; in this and other ways, their perspectives on research have been a significant influence on me.

I would like to thank my co-authors who have been a part of the research behind this thesis: Gerardo Adesso, John Calsamiglia, Animesh Datta, Paweł Horodecki, Julia Kempe,

Marco Piani, Jamie Sikora, Sarvagya Upadhyay, and Andreas Winter. It has been an honor working with and learning from you.

Over my time at Waterloo, I have been lucky enough to have had a circle of great friends. At some point it was decided that, having used the words "Hamiltonian" and "ground state energy" one time too many, that I had become a physicist, and a doodle of "photon Sev" mysteriously appeared on my office wall. Thank you for the great times, they will be sorely missed.

I am always grateful to my family, who has tirelessly supported and believed in me. Without their love and care, I would not and could not be where I am today.

Finally, words cannot express my gratitude to my wife, Mareike Müller. Together we lived in a "rabbit box" on campus for four years working on our Ph.D.'s. With any other person in such constantly close proximity, I think I would have lost my mind. But with her, it was a joy. Thank you for the wonderful experience, love, and support.

Financial support. I would like to thank the following agencies and programs for their funding support over the course of my Ph.D. studies: Natural Sciences and Engineering Research Council of Canada (NSERC), NSERC Michael Smith Foreign Study Supplement program, David R. Cheriton Scholarship program, EU-Canada Exchange program, the Institute for Quantum Computing at the University of Waterloo, and the Graduate Studies Office at the University of Waterloo.

The reader is referred to the end of each chapter for chapter-specific acknowledgements.

Dedication

To my family for their love and support, the foundation upon which all other success can be built.

Table of Contents

Chapter 1

Introduction

> *The "paradox" is only a conflict between reality and your feeling of what reality "ought to be."* — Richard Feynman, 1964 [95].

From its earliest days, the theory of quantum mechanics puzzled its inventors. In 1935, for example, Einstein, Podolsky, and Rosen published their now famous paper rejecting quantum mechanics as a complete physical theory [89]. The problem? The mathematical theory of quantum mechanics predicts certain physical phenomena which are completely at odds with our everyday understanding of the world around us. To put this into everyday language, in 1935 Schrödinger proposed [221] a thought experiment now known as *Schrödinger's cat*, in which under certain circumstances, a cat in a closed box is predicted by quantum mechanics to be both alive and dead, *at the same time*. What could this mean? And how much did it trouble the discoverers of quantum mechanics, if it led them to ask questions such as:

> *I recall that during one walk Einstein suddenly stopped, turned to me and asked whether I really believed that the moon exists only when I look at it.*
> — Abraham Pais [5].

Clearly, quantum mechanics was not an easy pill to swallow, even for the fathers of the theory, many of whom rejected their beautiful child at the time.

Fast forwarding to the end of the 20th century, however, physicists and computer scientists came to a startling realization: As strange as quantum mechanics may seem, if its peculiarities could somehow be computationally harnessed, then the possibility of

outperforming classical computers with so-called *quantum* computers may indeed exist. In 1982, for example, physicist Richard Feynman proposed [96] the notion of building a quantum computer in order to simulate physical quantum systems faster then apparently possible with a classical computer (see also Benioff [43, 44, 45]). On the computer science side, in 1985 David Deutsch demonstrated a quantum algorithm which outperformed the best possible classical deterministic algorithms for what is now referred to as *Deutsch's problem* [84]. Thus, the roots of the field of quantum computation were sown. Two and a half decades later, we now have a number of good reasons for seriously devoting research effort to the field of quantum computing, which we now discuss.

Relevance. We now state three reasons which, in our opinion, justify the study of quantum computation and information. The first is from an engineering-oriented perspective. Up until 2005, the speed of microprocessors increased rapidly, primarily through the brute force approach of increasing the number of transistors able to fit on a single microchip. Indeed, Intel's original Pentium P5 processor, released in 1993, had a clock speed of 60 MHz, and consisted of 3.1 million transistors [1]. By 2005, Intel's Pentium 4E Prescott processor was up to 3.8 GHz, and packed in a whopping 169 million transistors. Yet, in 2005, something curious happened: Intel introduced its first *dual-core* chip, the Pentium D Smithfield, which clocked in not at 3.8 GHz, but at a *slower* 3.2 Ghz. What happened? It turns out that the brute force approach to building faster processors has a number of seemingly fundamental problems, such as excess heat production and energy loss [3]; however, the primary problem of interest in this thesis is that at the scale current microchip components are approaching, the pertinent laws of physics are no longer those of classical mechanics, but rather those of *quantum* mechanics [2]. This raises the natural question: *Why not just build a computer which works based on the laws of quantum mechanics to begin with, i.e. a* quantum *computer?*

The second motivation for studying quantum computing, and perhaps the most commonly cited one, came with a startling discovery: Peter Shor's quantum factoring algorithm of 1994 [224]. As whether the question of whether factoring large integers can be done efficiently on a classical computer has long been open, Shor's algorithm is in itself arguably a strong indication that the quantum computational model is indeed one deserving of study. Further, since the algorithm's inception, a number of other instances of quantum speedup have been uncovered, from Grover's algorithm for unstructured search [122] (which yields a square root speedup for NP-complete problems over the brute force approach) to the evaluation of NAND trees [92, 24, 66] to estimating quantities related to solving systems of linear equations [129], among others.

The reasons stated thus far, however, are rather "selfish", aiming to exploit quantum

mechanics to serve the purpose of the computer science community. There is another view regarding the study of quantum computing which follows the converse mantra: *Ask not what quantum mechanics can do for you, but what you can do for quantum mechanics.* Indeed, as computation is inherently physical, it follows that understanding the limits of quantum computation yields new tools for studying the properties of quantum mechanics itself. A primary example of this, discussed further in Section 1.5.4, is that via quantum complexity theory, one can give a rigorous proof that a significant problem in quantum mechanics, that of estimating the ground state energy of a given local Hamiltonian, cannot be solved efficiently (modulo standard complexity theoretic conjectures). Thus, the third reason for studying quantum computation is that it not only allows us to learn about the limits of computing, but also of physics itself. Moreover, there has even been a *pedagogical* benefit to physics from quantum computing; apparently, there is a growing movement to replace the teaching of introductory quantum mechanics using, say, the model of the hydrogen atom, with the simpler model of quantum bits and quantum computation [27] (see Chapter Notes and History for Chapter 10 therein).

In closing, we have provided three motivations for studying quantum computing from engineering, computer science, and physics standpoints. In practice, however, it is of course not until a thorough study of quantum computing is undertaken that we will know the precise extent to which the field will prove relevant, particularly from a practical technological perspective. Such uncertainty lies unfortunately (or fortunately, for the adventurous type) at the very heart of the nature of our work as researchers. In the words of one of our greats:

> *If we knew what it was we were doing, it wouldn't be called 'research', would it?*
> — Albert Einstein [5].

Focus of this thesis. The field of quantum computation and information nowadays covers a broad expanse of topics, with research areas ranging from computer-science-motivated topics such as quantum algorithms and quantum proof systems, to engineering or experimental physics-oriented topics such as how to actually build a quantum computer in a lab, to theoretical-physics-motivated topics such as the limits of physical theories and the correlations between systems they allow. In this thesis, we focus on three particular areas of interest: Approximation of quantum problems, quantum proof systems, and quantum correlations. We briefly describe each area below. As each (research) chapter is intended to be as self-contained as possible, we defer more in-depth introductions to the beginning of each relevant chapter.

Our first area of interest is that of approximating quantum problems. Here, by a *quantum* problem, we are referring to a computational problem which is in some sense intrinsically related to physical quantum systems in nature. From a complexity theoretic perspective, we define such problems as those which are complete for quantum complexity classes. (Relevant quantum complexity classes are defined in Section 1.5.) In particular, the canonical quantum problem generalizing classical constraint satisfaction which we are interested in here is called the local Hamiltonian problem, and it is complete for a quantum generalization of NP. (This problem is important from both a quantum complexity theoretic and physics point of view, and as such is given a thorough treatment in Section 1.6.) The primary aim of our research in this area is to ask how well such problems can be *approximated rigorously*, in the well-studied classical sense of *approximation algorithms* and *hardness of approximation* [236]. In the quantum complexity theoretic setting, this approach to approximating physically relevant quantum problems is very much in its infancy, and it complements decades of effort by the physics community on similar problems using different tools involving heuristics (see e.g. [204] for a brief survey). Based on joint work with Julia Kempe, Chapters 2 and 3 discuss our results in this area, the first of which is a positive result regarding approximation algorithms for the local Hamiltonian problem, and the second of which is a negative result involving hardness of approximation for a new quantum complexity class generalizing the second level of the well-known polynomial-time hierarchy, Σ_2^p.

Our second area of interest deals with quantum proof systems. In the classical setting, proof systems are one of the cornerstones of complexity theory, with wide-ranging impact from the theory of NP-completeness [72, 179] to the stunning PCP theorem [30, 29] of the early 1990's. It is thus natural to consider studying *quantum* proof systems, beginning with a quantum generalization of NP called Quantum Merlin Arthur (QMA). However, just as quantum mechanics offers new quantum phenomena to be harnessed for the purpose of computation, such phenomena now play intriguing roles in quantum proof systems. In particular, their presence can turn trivial questions in the classical setting into highly non-trivial questions in the quantum setting. For example, in the classical setting, modifying NP to allow multiple provers is straightforwardly equivalent in expressive power to the original definition of NP, since a single prover can straightforwardly simulate multiple provers. However, the question of whether QMA with multiple provers is equal to QMA is very challenging, due to the possible presence of strong correlations between quantum systems known as *quantum entanglement*. In joint work with Jamie Sikora and Sarvagya Upadhyay, Chapter 4 studies variants of this stubbornly open question.

Our final area of interest is the study of quantum correlations. As mentioned when discussing quantum proof systems above, a pair of quantum systems can display very

strong correlations known as entanglement, which is a purely quantum phenomenon; such correlations are not possible in the classical setting. As a testament to the mysterious nature of quantum mechanics, however, after nearly a century of study, it has only been in recent years that a new type of purely quantum correlation has been identified, known simply as *non-classical* correlations. Some of the biggest questions in this area are how to quantify and provide operational interpretations for such correlations, as well as to understand whether and how they may be exploited for computational gain. In joint work with Animesh Datta, Chapter 5 studies the role of such correlations in quantum computation. Chapter 6 then proposes and studies a novel approach for quantifying such non-classical correlations. Finally, Chapters 7 (joint work with Marco Piani, Gerardo Adesso, John Calsamiglia, Paweł Horodecki, and Andreas Winter) and 8 (joint work with Marco Piani, Gerardo Adesso, John Calsamiglia, and Paweł Horodecki) introduce and study a new protocol which provides an operational interpretation for non-classical correlations by *activating* them into entanglement.

1.1 Organization

This thesis is organized as follows. In the remainder of this section, we provide background on the basics of quantum computation and information (Section 1.4), and follow with brief technical expositions of the various topics studied in this thesis: Quantum computational complexity theory (Section 1.5) and quantum entanglement and non-classical correlations (Section 1.6).

The remaining chapters are focused as follows. Chapters 2 and 3 study the approximability of quantum complexity theoretic problems, such as the local Hamiltonian problem and its variants. Specifically, Chapter 2 presents our approximation algorithm for the local Hamiltonian problem. Chapter 3 then introduces our quantum generalization of Σ_2^p, and shows completeness and hardness of approximation for it with respect to new local Hamiltonian-like quantum covering problems we define.

Chapter 4 discusses our results regarding multi-prover quantum proof systems, showing that in a certain setting, multiple quantum provers are no more powerful than a single prover.

Chapters 5, 6, 7, 8 discuss non-classical correlations in quantum systems beyond entanglement. Specifically, Chapter 5 first motivates this direction of work by studying models of quantum computing and communication where entanglement does not seem to explain the advantage gained in the quantum setting over classical computation. Chapter 6 then

presents a novel approach for quantifying non-classical correlations in quantum systems based on local unitary operations. Chapter 7 gives an operational interpretation to such non-classical correlations by demonstrating an explicit protocol through which such correlations can be "activated" into entanglement. Chapter 8 further studies and attempts to extend the framework of the activation protocol of Chapter 7.

We now begin in Section 1.2 by collecting common notation used throughout this thesis.

1.2 Notation

The following notation is assumed throughout this thesis. The symbols $\mathbb{C}$, $\mathbb{R}$, $\mathbb{Z}$, and $\mathbb{N}$ denote the sets of complex, real, integer, and natural numbers, respectively. For m a positive integer, the notation $[m]$ indicates the set $\{1, \ldots, m\}$. The terms $\mathcal{L}(\mathcal{X})$, $\mathcal{H}(\mathcal{X})$, $\mathrm{Pos}\,(\mathcal{X})$, and $\mathcal{D}(\mathcal{X})$ denote the sets of linear, Hermitian, positive semidefinite, and density operators acting on complex Euclidean space $\mathcal{X}$, respectively. The projector onto space $\mathcal{X}$ is denoted $\Pi_{\mathcal{X}}$. We sometimes use the shorthand $\mathcal{B} := \mathbb{C}^2$. The notation $A \succeq B$ means operator $A - B$ is positive semidefinite. The smallest (largest) eigenvalue of $A \in \mathcal{H}(\mathcal{X})$ is given by $\lambda_{\min}(A)$ ($\lambda_{\max}(A)$). The trace, Frobenius, and spectral (or operator) norms of $A \in \mathcal{L}(\mathcal{X})$ are defined as

$$\| A \|_{\mathrm{tr}} := \mathrm{Tr}\left(\sqrt{A^\dagger A}\right), \qquad \| A \|_{\mathrm{F}} := \sqrt{\mathrm{Tr}(A^\dagger A)}, \qquad \| A \|_\infty := \max_{|x\rangle \in \mathcal{X} \text{ s.t. } \| x \|_2 = 1} \| A|x\rangle \|_2 , \tag{1.1}$$

respectively, where $:=$ denotes a definition. The (m,n)th entry of matrix A is given by $A(m,n)$. We define the *encoding* or *description* of a matrix A as a classical description of the entries of A. Specifically, let $\langle \Delta \rangle$ denote the number of bits used to encode $\Delta \in \mathbb{C}$ to some desired precision. Then, we define the length of the encoding of A by $\langle A \rangle := \sum_{m,n} \langle A(m,n) \rangle$. We extend this straightforwardly to sums of matrices; for example, $\langle \sum_i A_i \rangle = \sum_i \langle A_i \rangle$. The notation $\mathbf{v}$ denotes a vector. Unless otherwise noted, all logarithms are taken to base two. We sometimes use the shorthand $\mathrm{poly}(n)$ to mean $p(n)$ for some fixed polynomial p.

1.3 Linear algebra

We now briefly review basic concepts from linear algebra crucial to the content of this thesis. Parts of this section follow the course notes of Watrous [246, 245]; the reader is

also referred to the text of Horn and Johnson [143] for further details. Those familiar with basic linear algebra can safely skim over this section or refer to it as needed.

Complex Euclidean spaces. The setting in which all the excitement takes place is that of a complex Euclidean space $\mathcal{X}$, defined as follows. Let Σ be a finite, non-empty set. Consider the set of all functions from Σ to the complex numbers $\mathbb{C}$, denoted $\mathbb{C}^\Sigma$. Then, define for any $\mathbf{u}, \mathbf{v} \in \mathbb{C}^\Sigma$ and $\alpha \in \mathbb{C}$ the addition and scalar multiplication operations in the standard way: The addition $\mathbf{u} + \mathbf{v} \in \mathbb{C}^\Sigma$ obeys $(\mathbf{u} + \mathbf{v})(i) = \mathbf{u}(i) + \mathbf{v}(i)$ for all $i \in \Sigma$, and scalar multiplication $\alpha\mathbf{u} \in \mathbb{C}^\Sigma$ obeys $(\alpha\mathbf{u})(i) = \alpha\mathbf{u}(i)$ for all $i \in \Sigma$. Then, the set $\mathbb{C}^\Sigma$ along with these operations is known as a complex Euclidean space, which we denote as $\mathcal{X}$. The dimension of $\mathcal{X}$ is given by $|\Sigma|$, the cardinality of Σ. For concreteness, we henceforth assume $\Sigma = [d]$ for $[d] := \{1, 2, \ldots, d\}$, and use the simplified notation $\mathbb{C}^\Sigma = \mathbb{C}^d$.

We think of (column) vectors $\mathbf{v} \in \mathcal{X}$ as d-tuples, i.e.

$$\mathbf{v} = \begin{pmatrix} v(1) \\ \vdots \\ v(d) \end{pmatrix} \tag{1.2}$$

for $v(i) \in \mathbb{C}$. In quantum computation, $\mathbf{v}$ is commonly denoted using $|v\rangle$. Here, $|\cdot\rangle$ is called Dirac notation, also sometimes affectionately known as "dog-houses" for vectors [243]. A remark about vector notation: Generally, our choice of notation $\mathbf{v}$ or $|v\rangle$ will be dictated by context. For example, when a vector is to be interpreted as a quantum state, we shall use Dirac notation $|v\rangle$; otherwise, we typically revert to the notation $\mathbf{v}$. An exception to this rule, even in purely linearly algebraic contexts, is when it is more convenient to use Dirac notation, such as when vectors are to be labeled by complicated expressions. In much of the introductory discussion on linear algebra that follows, we assume $\mathbf{v} = |v\rangle$ holds for the pedagogic purpose of familiarizing the reader with Dirac notation. However, in general this equality is not assumed to hold; for example, the zero vector $\mathbf{0}$ is not equal to $|0\rangle = (1, 0)^T$. We hope the distinction will be clear from context.

Continuing, the conjugate transpose of $\mathbf{v}$ is denoted $\mathbf{v}^\dagger$, or $\langle v|$ in Dirac notation, and is the row vector

$$\mathbf{v}^\dagger = \langle v| = \left(\overline{v(1)}, \overline{v(2)}, \ldots, \overline{v(d)}\right), \tag{1.3}$$

for $\overline{a}$ the complex conjugate of $a \in \mathbb{C}$.

Vector norms. For any two vectors $\mathbf{v}, \mathbf{w} \in \mathcal{X}$, we define their inner product as

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^\dagger \mathbf{w} = \langle v | w \rangle = \sum_{i=1}^{d} \overline{v(i)} w(i). \tag{1.4}$$

Then, we measure the length of $\mathbf{v} \in \mathbb{C}^d$ via the *Euclidean norm*, defined as $\| \mathbf{v} \|_2 = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$. The Euclidean norm is just one of an entire class of norms known as p-norms, defined for $p \in [1, \infty)$ such that

$$\| \mathbf{v} \|_p := \left(\sum_{i=1}^{d} |v(i)|^p \right)^{\frac{1}{p}}, \tag{1.5}$$

and for $p = \infty$ as $\| \mathbf{v} \|_\infty := \left(\max_{i \in [d]} |v(i)| \right)$. Note that setting $p = 2$ yields the Euclidean norm. The p-norms have the following properties:

1. (Positive scalability) $\| a\mathbf{v} \|_p = |a| \, \| \mathbf{v} \|_p$ for $a \in \mathbb{C}$.
2. (Triangle inequality) For any $\mathbf{v}, \mathbf{w} \in \mathcal{X}$, $\| \mathbf{v} + \mathbf{w} \|_p \leq \| \mathbf{v} \|_p + \| \mathbf{w} \|_q$.
3. For $\mathbf{v} \in \mathcal{X}$, if $\| \mathbf{v} \| = 0$, then $\mathbf{v} = \mathbf{0}$, where $\mathbf{0}$ denotes the zero vector whose entries are all zero.

From the first two properties, we conclude that for all $\mathbf{v} \in \mathcal{X}$, $\| \mathbf{v} \|_p \geq 0$, since

$$0 = |0| \, \| \mathbf{0} \|_p = \| \, 0 \cdot \mathbf{0} \|_p = \| \mathbf{0} \|_p = \| \mathbf{v} - \mathbf{v} \|_p \leq \| \mathbf{v} \|_p + \| -\mathbf{v} \|_p \leq 2 \| \mathbf{v} \|_p . \tag{1.6}$$

A useful inequality regarding inner products is the Hölder inequality, which states that for any $\mathbf{v}, \mathbf{w} \in \mathcal{X}$,

$$|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \| \mathbf{v} \|_p \, \| \mathbf{w} \|_q \tag{1.7}$$

for $\frac{1}{p} + \frac{1}{q} = 1$. (For $p = 1$, $q = \infty$.) When $p = q = 2$, we recover the Cauchy-Schwarz inequality. As a testament to the applicability of the latter, we show that $\| v \|_1 \leq \sqrt{d} \, \| v \|_2$, a frequently useful inequality. Let $\mathbf{j}$ be the d-dimensional all-ones vector and $|\mathbf{v}|$ the entry-wise absolute value of $\mathbf{v}$. Then:

$$\| v \|_1 = \langle \mathbf{j}, |\mathbf{v}| \rangle \leq |\langle \mathbf{j}, |\mathbf{v}| \rangle| \leq \| \mathbf{j} \|_2 \, \| \, |\mathbf{v}| \, \|_2 = \sqrt{d} \, \| \mathbf{v} \|_2 . \tag{1.8}$$

It also holds that $\| \mathbf{v} \|_2 \leq \sqrt{d} \, \| \mathbf{v} \|_\infty$, and conversely that $\| \mathbf{v} \|_1 \geq \| \mathbf{v} \|_2 \geq \| \mathbf{v} \|_\infty$.

Orthonormal bases. A set of vectors $\{\mathbf{v}_i\} \subseteq \mathcal{X}$ is *orthogonal* if for all $i \neq j$, $\langle \mathbf{v}_i, \mathbf{w}_j \rangle = 0$, and *orthonormal* if $\langle \mathbf{v}_i, \mathbf{w}_j \rangle = \delta_{ij}$. Here, δ_{ij} is the Kroenecker delta, whose value is 1 if $i = j$ and 0 otherwise. Every complex Euclidean space $\mathcal{X}$ of dimension d has an orthonormal *basis* consisting of d elements, where a basis is a set of vectors $\{\mathbf{v}_i\} \subseteq \mathcal{X}$ such that any $\mathbf{w} \in \mathcal{X}$ can be expressed as

$$\mathbf{w} = \sum_{i=1}^{d} \alpha_i \mathbf{v}_i \tag{1.9}$$

for some $\{\alpha_i\} \subseteq \mathbb{C}$. A common basis for $\mathcal{X}$ is the *computational* or *standard* basis $\{\mathbf{e}_i\}$, defined such that $\mathbf{e}_i(j) = \delta_{ij}$. In Dirac notation, we frequently denote this basis simply as $\{|i\rangle\}_{i=1}^{d}$.

Linear operators and matrices. Given two complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, a *linear operator* or *linear map* from $\mathcal{X}$ to $\mathcal{Y}$ is a map $\Phi : \mathcal{X} \mapsto \mathcal{Y}$ with the property that

$$\Phi\left(\sum_i \alpha_i \mathbf{v}_i\right) = \sum_i \alpha_i \Phi(\mathbf{v}_i), \tag{1.10}$$

where $\{\mathbf{v_i}\} \subseteq \mathcal{X}$. The set of all such linear maps from $\mathcal{X}$ to $\mathcal{Y}$ is denoted $\mathcal{L}(\mathcal{X}, \mathcal{Y})$, which when coupled with operations for addition and scalar multiplication in the standard way, yields a vector space of dimension $\dim(\mathcal{X})\dim(\mathcal{Y})$. Here, $\dim(\mathcal{X})$ is the dimension of $\mathcal{X}$. For brevity, we use the shorthand $\mathcal{L}(\mathcal{X})$ to mean $\mathcal{L}(\mathcal{X}, \mathcal{X})$.

A convenient way to represent and study linear maps is via their matrix representation. Here, an $m \times n$ *matrix* A is a two-dimensional array of complex numbers whose (i, j)th entry is denoted $A(i, j) \in \mathbb{C}$ for $i \in [m]$, $j \in [n]$. To represent a linear map $\Phi : \mathbb{C}^n \mapsto \mathbb{C}^m$ as an $m \times n$ matrix A_Φ, recall that the action of a map is completely specified by its action on a basis. Specifically, the ith column of A_Φ is given by $\Phi(\mathbf{e}_i)$ for $\{\mathbf{e}_i\}$ the standard basis for $\mathbb{C}^n$, or

$$A_\Phi = \left[\ \Phi(\mathbf{e}_1), \Phi(\mathbf{e}_2), \ldots, \Phi(\mathbf{e}_m)\ \right]. \tag{1.11}$$

Recovering Φ from A_Φ thus also follows immediately from this view. When we henceforth discuss $A \in \mathcal{L}(\mathcal{X})$, we are implicitly referring to the matrix representation of map A.

The product AB of two $d \times d$ matrices A and B is defined such that

$$AB(i, j) = \langle \overline{\mathbf{r}}_i^A, \mathbf{c}_j^B \rangle \tag{1.12}$$

for $\mathbf{r}_i^A$ the ith row of A and $\mathbf{c}_j^B$ the jth column of B. In general, it is not true that $AB = BA$. The difference $AB - BA$ is called the *commutator* $[A, B]$ of A and B, and the *anti-commutator* is $\{A, B\} = AB + BA$.

The *rank* of $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ is the dimension of its *image*, where the latter is defined as $\mathrm{Im}(A) := \{\mathbf{y} \in \mathcal{Y} \mid \mathbf{y} = A\mathbf{x} \text{ for some } \mathbf{x} \in \mathcal{X}\}$. The rank satisfies

$$\mathrm{rank}(AB) \leq \min\{\mathrm{rank}(A), \mathrm{rank}(B)\}. \tag{1.13}$$

Defining the *null space* or *kernel* of $A \in \mathcal{L}(\mathcal{X})$ as $\mathrm{Ker}(A) := \{\mathbf{v} \in \mathcal{X} \mid A\mathbf{v} = 0\}$, it holds that $\dim(\mathrm{Ker}(A)) + \dim(\mathrm{Im}(A)) = d$.

Eigenvalues and eigenvectors. For any $A \in \mathcal{L}(\mathcal{X})$, we say $\mathbf{v}$ is an *eigenvector* of A with *eigenvalue* λ if $\mathbf{v} \neq \mathbf{0}$ and $A\mathbf{v} = \lambda A$. The multiset of eigenvalues of A (with multiplicity) is known as its *spectrum*. The eigenvalues of A arise as the roots of the degree-d *characteristic polynomial* of A, p_A, defined such that

$$p_A(x) := \det(xI - A), \tag{1.14}$$

where $I(i, j) := \delta_{ij}$ is the *Identity* matrix and det is the *determinant*. One way to define the latter, known as the Laplace expansion, is via the recursive definition

$$\det(A) = \sum_{j=1}^{d} (-1)^{i+j} A(i, j) \det(A_{ij}). \tag{1.15}$$

Here, A_{ij} is the matrix obtained from A by deleting row i and column j, and we define the base case of this recursion (i.e. a 1×1 matrix $[c]$) as $\det([c]) = c$. This equation holds for any $i \in [d]$.

Matrix operations. A number of operations on matrices $A \in \mathcal{X}$ arise repeatedly in quantum computing. First, the *complex conjugate*, *transpose* and *adjoint* operations are respectively defined via

$$\overline{A}(i, j) := \overline{(A(i, j))} \qquad A^T(i, j) := A(j, i) \qquad A^\dagger := (\overline{A})^T. \tag{1.16}$$

These operations apply to vectors as well so that $\langle v|$, defined in Equation (1.3), is simply $|v\rangle^\dagger$.

The *trace* of A is a linear function defined as $\mathrm{Tr}(A) := \sum_{i=1}^{d} A(i, i) = \sum_{i=1}^{d} \lambda_i(A)$, where $\{\lambda_i(A)\} \subseteq \mathbb{C}$ are the eigenvalues of A. Henceforth, when clear from context, we

simply write λ_i for the latter. The trace has the useful property of being *cyclic*, i.e. $\mathrm{Tr}(ABC) = \mathrm{Tr}(CAB)$. With the trace in hand, we can define an inner product on $\mathcal{L}(\mathcal{X})$ as $\langle A, B\rangle = \mathrm{Tr}(A^\dagger B)$.

The *tensor product* is an important operation through which joint quantum systems can be described. Specifically, for complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, their tensor product is $\mathcal{X} \otimes \mathcal{Y} = \mathbb{C}^{d_x \times d_y}$. For vectors $\mathbf{u} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{Y}$, we define for all $i \in [d_x]$ and $j \in [d_y]$

$$(\mathbf{u} \otimes \mathbf{v})(i, j) := u(i)v(j). \tag{1.17}$$

For linear operators $A \in \mathcal{L}(\mathcal{X})$, $B \in \mathcal{L}(\mathcal{Y})$, $A \otimes B$ yields a complex matrix whose index sets are given by $([d_x] \times [d_y], [d_x] \times [d_y])$, such that

$$(A \otimes B)((i_1, j_1), (i_2, j_2)) := A(i_1, i_2)B(j_1, j_2) \tag{1.18}$$

for all $i_1, i_2 \in [d_x]$ and $j_1, j_2 \in [d_y]$. The tensor product has the following properties for any $A, C \in \mathcal{X}$, $B, D \in \mathcal{Y}$, $c \in \mathbb{C}$:

$$\begin{aligned}
(A + C) \otimes B &= A \otimes B + C \otimes B && (1.19)\\
A \otimes (B + D) &= A \otimes B + A \otimes D && (1.20)\\
c(A \otimes B) &= (cA) \otimes B = A \otimes (cB) && (1.21)\\
(A \otimes B)(C \otimes D) &= AC \otimes BD && (1.22)\\
\mathrm{Tr}(A \otimes B) &= \mathrm{Tr}(A)\mathrm{Tr}(B) && (1.23)\\
(A \otimes B)^\dagger &= A^\dagger \otimes B^\dagger. && (1.24)
\end{aligned}$$

These properties hold analogously in the vector setting.

Given the composition of two spaces $\mathcal{X}$ and $\mathcal{Y}$ via the tensor product, we also require an operation in the reverse direction for removing one of these spaces. For this, we define the linear *partial trace* map. Specifically, for $A \otimes B \in \mathcal{L}(\mathcal{X} \otimes \mathcal{Y})$, the partial trace $\mathrm{Tr}_{\mathcal{X}}(A \otimes B) \in \mathcal{Y}$ is defined as

$$\mathrm{Tr}_{\mathcal{X}}(A \otimes B) := \mathrm{Tr}(A)B. \tag{1.25}$$

Alternatively, for any orthonormal basis $\{\mathbf{v}_i\}_{i=1}^d$ for $\mathcal{X}$, we can write for $A \in \mathcal{L}(\mathcal{X} \otimes \mathcal{Y})$

$$\mathrm{Tr}_{\mathcal{X}}(A) = \sum_{i=1}^{d} \left(\mathbf{v}_i^\dagger \otimes I\right) A \left(\mathbf{v}_i \otimes I\right). \tag{1.26}$$

Special classes of operators. A few classes of linear operators play important roles in quantum computing. The first of these is the class of *Hermitian* operators $\mathcal{H}(\mathcal{X}) \subseteq \mathcal{L}(\mathcal{X})$, defined as the set of $A \in \mathcal{L}(\mathcal{X})$ satisfying $A^\dagger = A$. As the set of Hermitian operators is closed under addition and real scalar multiplication, and since $\langle A, B\rangle \in \mathbb{R}$ for all $A, B \in \mathcal{H}(\mathcal{X})$, it follows that $\mathcal{H}(\mathcal{X})$ forms a real inner product space of dimension d^2.

The eigenvalues of Hermitian operators are real. If the eigenvalues of Hermitian A are in $\{0, 1\}$, then equivalently $A^2 = A$, and A is called an (orthogonal) projection. (Non-Hermitian A satisfying $A^2 = A$ are called *oblique* projections, and are not used here.)

More generally, a Hermitian matrix $A \in \mathcal{H}(\mathcal{X})$ whose eigenvalues are all non-negative is called *positive semidefinite*, denoted $A \succeq 0$ (more generally, the notation $A \succeq B$ means $A - B \succeq 0$). Positive semidefinite matrices $A \in \mathcal{H}(\mathcal{X})$ can equivalently be characterized as follows:

- $\mathbf{x}^\dagger A \mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathcal{X}$.
- $A = B^\dagger B$ for some $B \in \mathcal{L}(\mathcal{X})$.

The set of positive semidefinite operators acting on $\mathcal{X}$ is denoted $\mathrm{Pos}(\mathcal{X})$.

Next, a *unitary* operator $U \in \mathcal{U}(\mathcal{X})$ is defined as satisfying $UU^\dagger = U^\dagger U = I$. The eigenvalues of U are complex numbers of modulus 1. All unitary operators preserve the length of any vector $\mathbf{v}$, i.e. $\langle U\mathbf{v}, U\mathbf{v}\rangle = \langle \mathbf{v}, \mathbf{v}\rangle$. More generally, any $U \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ with $U^\dagger U = I_\mathcal{X}$ is called an *isometry*.

Hermitian, positive semidefinite, and unitary matrices are in fact all special cases of *normal* matrices A, defined such that $AA^\dagger = A^\dagger A$. Normal matrices are important due to the Spectral Decomposition theorem, which we discuss next.

Matrix decompositions. An extremely useful property of normal matrices A acting on $\mathcal{X}$ is that they can be written in terms of their *spectral decomposition*, i.e.

$$A = \sum_{i=1}^{d} \lambda_i |\lambda_i\rangle\langle\lambda_i| = UDU^\dagger, \tag{1.27}$$

where recall λ_i are the eigenvalues of A, the set $\{|\lambda_i\rangle\}_{i=1}^d$ is a corresponding orthonormal set of eigenvectors of A, $D = \mathrm{diag}(\{\lambda_i\})$ is a diagonal operator with entries $D(i, i) = \lambda_i$, and U is a unitary matrix whose ith column is $|\lambda_i\rangle$. Here we have switched to Dirac notation to highlight, in our opinion, one of its strengths — the ability to label vectors easily by

complicated expressions. Note that if $\lambda_i \neq \lambda_j$ for all i, j, then the set of eigenvectors above is unique.

A common problem in quantum mechanics is to analyze the spectrum of a sum of two matrices $A, B \in \mathcal{L}(\mathcal{X})$. In general, this is a difficult problem. However, if the matrices are normal and they *commute*, i.e. $[A, B] = 0$, then this task is made easier by the fact that A and B must *simultaneously diagonalize*. In other words for normal A and B, $[A, B] = 0$ if and only if there exists an orthonormal basis $\{|b_i\rangle\} \subseteq \mathcal{X}$ such that

$$A = \sum_{i=1}^{d} \lambda_i(A)|b_i\rangle\langle b_i|, \qquad B = \sum_{i=1}^{d} \lambda_i(B)|b_i\rangle\langle b_i|. \tag{1.28}$$

While the spectral decomposition holds only for normal matrices, a more general decomposition known as the *singular value* decomposition exists even for non-square matrices. The latter says that for any $d_y \times d_x$ matrix $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$, we have

$$A = UDV^\dagger \tag{1.29}$$

for $d_y \times d_y$ unitary U, $d_x \times d_x$ unitary V, and $d_y \times d_x$ diagonal matrix D whose entries $D(i, i)$ are non-negative real numbers called the *singular values* of A.

Operator functions. With the spectral decomposition in hand, we can now apply functions $f : \mathbb{C} \mapsto \mathbb{C}$ to normal operators $A \in \mathcal{X}$ as follows. Let A have spectral decomposition $A = \sum_{i=1}^{d} \lambda_i|\lambda_i\rangle\langle\lambda_i|$. Then, assuming $\{\lambda_i\}$ is a subset of the domain of f,

$$f(A) := \sum_{i=1}^{d} f(\lambda_i)|\lambda_i\rangle\langle\lambda_i|. \tag{1.30}$$

Three common functions f encountered in this thesis are $f(x) = e^x$, $f(x) = \log x$, and $f(x) = \sqrt{x}$, the operator functions of which are denoted as e^A, $\log A$, and $\sqrt{A}$, respectively. Here, the logarithm is taken to base two.

Operator norms. Similar to the p-norms we defined for vectors, a useful class of norms for measuring the "length" or "magnitude" of a matrix are the Schatten p-norms. Their definition is simple: For any $p \in [1, \infty]$, let $\sigma(A)$ denote the vector of singular values of $A \in \mathcal{X}$. Then,

$$\| A \|_p := \| \sigma(A) \|_p . \tag{1.31}$$

A particularly nice aspect of this definition is that for Hermitian operators, $\sigma_i(A) = |\lambda_i(A)|$. Moreover, properties of the vector p-norms carry over straightforwardly to the Schatten p-norms, such as the Hölder inequality, positive scalability, and the triangle inequality.

Some further important properties of the p-norms for any $A \in \mathcal{L}(\mathcal{X})$ are:

1. $\| A \|_p = \left\| \overline{A} \right\|_p = \left\| A^T \right\|_p$, from which also $\| A \|_p = \left\| A^\dagger \right\|_p$.
2. (Invariance under isometries) $\left\| UAV^\dagger \right\|_p = \| A \|_p$ for any isometries U and V for which $UAV^\dagger$ is well-defined.
3. $\| ABC \|_p \leq \| A \|_\infty \| B \|_p \| C \|_\infty$.
4. (Submultiplicativity) $\| AB \|_p \leq \| A \|_p \| B \|_p$. This follows from Property 3.

There are three specific values of p of interest here: $p = 1$, $p = 2$, and $p = \infty$. They correspond to the *trace*, *Frobenius*, and *spectral* (or *operator*) norms, respectively, and can alternatively be defined as

$$\| A \|_{\mathrm{tr}} := \mathrm{Tr}\left(\sqrt{A^\dagger A}\right), \qquad \| A \|_{\mathrm{F}} := \sqrt{\mathrm{Tr}(A^\dagger A)}, \qquad \| A \|_\infty := \max_{|x\rangle \in \mathcal{X} \text{ s.t. } \| x \|_2 = 1} \| A|x\rangle \|_2 . \tag{1.32}$$

The trace norm has two further properties of interest: First, it is non-increasing under the partial trace, meaning that for $A \in \mathcal{L}(\mathcal{X} \otimes \mathcal{Y})$, $\| \mathrm{Tr}_{\mathcal{Y}}(A) \|_{\mathrm{tr}} \leq \| A \|_{\mathrm{tr}}$. Second, for unit vectors $\mathbf{u}, \mathbf{v} \in \mathcal{X}$ we have

$$\left\| \mathbf{u}\mathbf{u}^\dagger - \mathbf{v}\mathbf{v}^\dagger \right\|_{\mathrm{tr}} = 2\sqrt{1 - |\langle \mathbf{u}, \mathbf{v} \rangle|^2} \leq 2 \| \mathbf{u} - \mathbf{v} \|_2 . \tag{1.33}$$

The second inequality follows by expanding the definition of the Euclidean norm and applying the identity $1 - x^2 \leq 2(1 - x)$. The first equality follows [246] by noting that $A := \mathbf{u}\mathbf{u}^\dagger - \mathbf{v}\mathbf{v}^\dagger$ is Hermitian, and so its trace norm is a function of the absolute values of its eigenvalues, which we now analyze. Since $\mathrm{rank}(A) \leq 2$ and $\mathrm{Tr}(A) = 0$, its spectrum must be $\{\lambda, -\lambda, 0, \ldots, 0\}$ for some $\lambda \in \mathbb{R}$. Thus, $\mathrm{Tr}(A^2) = 2\lambda^2$. However, a direct evaluation of $\mathrm{Tr}(A^2)$ from the definition of A also reveals $\mathrm{Tr}(A^2) = 2 - 2\,|\langle \mathbf{u}, \mathbf{v} \rangle|^2$. Combining these two expressions for $\mathrm{Tr}(A^2)$, the claim follows.

Linear super-operators. We have discussed (linear) operators $\Phi : \mathcal{X} \mapsto \mathcal{X}$ and $\Phi : \mathcal{Y} \mapsto \mathcal{Y}$. Moving a step up the ladder, we can also discuss linear operators $\Phi : \mathcal{L}(\mathcal{X}) \mapsto \mathcal{L}(\mathcal{Y})$. Such maps are called linear *super-operators*. Bestowed with the standard definitions of

addition and scalar multiplication, the set of super-operators, denoted $T(\mathcal{X}, \mathcal{Y})$, forms a linear space. The tensor product operation applies analogously to super-operators as it did to operators.

The *adjoint* of super-operator $\Phi \in T(\mathcal{X}, \mathcal{Y})$, $\Phi^* \in T(\mathcal{Y}, \mathcal{X})$, is uniquely defined by the equation

$$\langle A, \Phi(B) \rangle = \langle \Phi^*(A), B \rangle, \tag{1.34}$$

which holds for all $B \in \mathcal{X}$ and $A \in \mathcal{Y}$.

Special classes of super-operators. From a quantum computing perspective, we are most interested in super-operators which are trace-preserving and completely positive (TPCP). A *trace-preserving* super-operator $\Phi \in T(\mathcal{X}, \mathcal{Y})$ is defined as satisfying

$$\mathrm{Tr}(A) = \mathrm{Tr}(\Phi(A)) \tag{1.35}$$

for any $A \in \mathcal{L}(\mathcal{X})$. To define a completely positive map, we first define a *positive* map $\Phi \in T(\mathcal{X}, \mathcal{Y})$ as satisfying $\Phi(A) \succeq 0$ for any $A \in \mathcal{L}(\mathcal{X})$ such that $A \succeq 0$. Then, a map $\Phi \in T(\mathcal{X}, \mathcal{Y})$ is called *completely positive* if $I_{\mathcal{L}(\mathcal{X})} \otimes \Phi$ is a positive map. Intuitively, a completely positive map Φ sends positive semidefinite operators to positive semidefinite operators, even if Φ acts on only part of a larger composite system.

Matrix representations of super-operators. Just as we discussed a matrix representation for linear operators, there are a number of useful matrix representations for linear super-operators. (See the notes of Watrous [247] for an excellent exposition.) Here, we discuss two particular representations used in this thesis, known as the *Stinespring* and *Kraus* representations.

The Strinespring representation lends a nice interpretation to admissible quantum maps later. Specifically, it says that the action of any TPCP map $\Phi \in T(\mathcal{X}, \mathcal{Y})$ on arbitrary $X \in \mathcal{L}(\mathcal{X})$ can be written as

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}(AXA^\dagger), \tag{1.36}$$

for some complex Euclidean space $\mathcal{Z}$ and some linear isometry $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$. Moreover, $\dim(\mathcal{Z})$ can be taken as $\dim(\mathcal{Z}) \leq \dim(\mathcal{X}) \dim(\mathcal{Y})$. In the context of quantum computation, it will be particularly useful to note that this is equivalent [21] to saying $\Phi(X)$ can be written as, for $\mathcal{Y} = \mathcal{Y}_1 = \mathcal{Y}_2$,

$$\Phi(X) = \mathrm{Tr}_{\mathcal{X} \otimes \mathcal{Y}_2} \left[U(X_{\mathcal{X}} \otimes |0\rangle\langle 0|_{\mathcal{Y}_1 \otimes \mathcal{Y}_2}) U^\dagger \right], \tag{1.37}$$

for some unitary $U \in \mathcal{U}(\mathcal{X} \otimes \mathcal{Y}_1 \otimes \mathcal{Y}_2)$.

We now define the Kraus representation, which is sometimes also known as the *operator-sum* representation [200]. The Kraus representation says that any TPCP map $\Phi \in T(\mathcal{X}, \mathcal{Y})$ can be expressed in terms of a set of *Kraus operators* $\{K_i\}_{i=1}^k \subseteq \mathcal{L}(\mathcal{X}, \mathcal{Y})$ such that

$$\Phi(X) = \sum_{i=1}^{k} K_i X K_i^\dagger, \tag{1.38}$$

where $\sum_{i=1}^k K_i^\dagger K_i = I_\mathcal{X}$ and $k \leq \dim(\mathcal{X}) \dim(\mathcal{Y})$.

1.4 Basics of quantum computation

We now introduce the basics of quantum computation. For further details, the interested reader is referred to the texts of Nielsen and Chuang [200], Kitaev, Shen, and Vyalyi [171], and Kaye, Laflamme, and Mosca [162]. From a computer scientist's perspective, note that the primary background required is *not* quantum physics, but rather linear algebra [143]. This is because, just as with any (say) sports game, in order to play the game, you simply have to learn the *rules* of the game. Quantum mechanics, in particular, has four simple rules, and they are all based on linear algebra. These rules govern the following four intuitively logical concepts: How a quantum state is described, how does one "read" or measure a quantum state, what operations can be performed on a quantum state, and finally, how does one describe multiple quantum systems jointly.

1.4.1 Describing quantum states

Let $\mathcal{X}$ denote a complex Euclidean space. Then, in a nutshell, any $\rho \in \mathrm{Pos}(\mathcal{X})$ with trace 1 describes a valid quantum state. Let us now provide some intuition as to how this statement comes about.

In classical computing, the basic unit of information is a bit, which takes on values in the set $\{0, 1\}$. One can equivalently encode a bit using the set $\{|0\rangle, |1\rangle\}$, where $\{|0\rangle, |1\rangle\} \subseteq \mathbb{C}^2$ is the standard basis for $\mathbb{C}^2$, i.e. $|0\rangle = (1, 0)^T$ and $|1\rangle = (0, 1)^T$. The key difference between classical bits and qubits is that in the quantum world, one can interpolate between the two discrete values $|0\rangle$ and $|1\rangle$ by taking a *superposition*, i.e. the vector

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{1.39}$$

describes a valid quantum state if $|\alpha|^2 + |\beta|^2 = 1$. In other words, any unit vector in $\mathbb{C}^2$ describes a quantum bit, or *qubit*.

More generally, assume $\mathcal{X}$ has dimension d. Then, any unit vector $|\psi\rangle \in \mathcal{X}$ describes a d-dimensional quantum state, sometimes dubbed a qu$d$it. Such vectors are called *pure* states, and do not yet capture the set of all possible d-dimensional quantum states. To complete the picture, we simply allow probabilistic mixtures of such pure states, more generally referred to as *mixed* states. Such probabilistic mixtures are described in the following straightforward manner, known as the *density matrix* formalism.

Associated with any probabilistic mixture is an *ensemble*,

$$\left\{\{p_i\}_{i=1}^k, \{|\psi_i\rangle\langle\psi_i|\}_{i=1}^k\right\}, \tag{1.40}$$

where $\{p_i\}_{i=1}^k$ forms a probability distribution and $\{|\psi_i\rangle\} \subseteq \mathcal{X}$ is a set of unit vectors. The corresponding mixed quantum state ρ is thus:

$$\rho = \sum_{i=1}^k p_i |\psi_i\rangle\langle\psi_i|. \tag{1.41}$$

Here, ρ is called the *density matrix* describing the underlying quantum state. We denote the set of density operators acting on $\mathcal{X}$ as $\mathcal{D}(\mathcal{X})$.

Let us now tie this back into the statement made at the beginning of this subsection. Note that since in Equation (1.41), ρ is a non-negative sum of positive semidefinite operators, we must have $\rho \succeq 0$. Moreover, by applying the cyclic property of the trace, we have $\mathrm{Tr}(\rho) = 1$, as claimed. Indeed, based on the exposition above, we can now intuitively see why any $\rho \in \mathcal{X}$ with $\rho \succeq 0$ and $\mathrm{Tr}(\rho) = 1$ describes a valid quantum state — simply take the spectral decomposition of ρ to recover an ensemble $\left\{\{p_i\}_{i=1}^k, \{|\psi_i\rangle\langle\psi_i|\}_{i=1}^k\right\}$.

We remark that although here we have attempted to present a simple exposition of how quantum states are classically described, in reality the precise interpretation of what such a classical description means is highly non-trivial and continues to be debated after decades of research.

1.4.2 Measuring quantum states

Now that we have a mathematical description of quantum states, we require a formalism for modeling how a quantum state is "observed", or measured. For this, let $\rho \in \mathcal{D}(\mathcal{X})$

be a density matrix. Then, a quantum *measurement* is formalized by a set of operators $\Pi := \{M_i\} \subseteq \mathcal{L}(\mathcal{X})$ satisfying

$$\sum_i M_i^\dagger M_i = I, \tag{1.42}$$

where the latter is called the *completeness relation*. The act of measuring ρ with Π is in general an inherently probabilistic process, *even if* ρ corresponds to a pure state (unlike in the classical case of bits). Specifically, when measuring ρ with respect to Π, we obtain outcome i with probability given by

$$\Pr(\text{outcome } i|\rho) = \mathrm{Tr}(M_i \rho M_i^\dagger). \tag{1.43}$$

Once a particular outcome i is observed, the state ρ "collapses" to a new state ρ' consistent with this outcome, i.e.

$$\rho' = \frac{M_i \rho M_i^\dagger}{\Pr(\text{outcome } i|\rho)}. \tag{1.44}$$

Note that the denominator above serves the role of renormalizing ρ' so that $\mathrm{Tr}(\rho') = 1$.

We have thus far described general measurements. Often, we are interested in the special case when each M_i is an orthogonal projection operator (not necessarily of rank one), such that $M_i M_j = \delta_{ij} M_i$. Such measurements are called *projective* or *von Neumann* measurements. A common way to represent a projective measurement is via an *observable* $M \in \mathcal{H}(\mathcal{X})$. Via the spectral decomposition, we can write $M = \sum_i \lambda_i \Pi_i$, where $\lambda_i \neq \lambda_j$ for $i \neq j$ and each Π_i is a projection operator (of rank possibly greater than one). Then, each eigenvalue λ_i corresponds to a distinct label for a measurement outcome, and the measurement operators are $M_i = \Pi_i$. An advantage of using observables is that the expected value of the measurement, denoted $\mathbb{E}_M$, takes a very simple form:

$$\mathbb{E}_M(\rho) = \sum_i \lambda_i \Pr(\text{outcome } i|\rho) = \sum_i \lambda_i \mathrm{Tr}(\Pi_i \rho \Pi_i^\dagger) = \sum_i \lambda_i \mathrm{Tr}(\Pi_i \rho) = \mathrm{Tr}(M\rho). \tag{1.45}$$

Finally, note that the framework above for general measurements $\Pi = \{M_i\}$ allows one to determine both the probability of outcome i, as well as the output state of the measurement process once i is read. If we only care about the former, as is the case in situations where the quantum system is only to be measured once and subsequently discarded, then this formalism is often simplified by defining positive semidefinite $E_i := M_i^\dagger M_i$ with $\sum_i E_i = I$. We hence have:

$$\Pr(\text{outcome } i|\rho) = \mathrm{Tr}(M_i \rho M_i^\dagger) = \mathrm{Tr}(M_i^\dagger M_i \rho) = \mathrm{Tr}(E_i \rho). \tag{1.46}$$

The set $\{E_i\}$ is called a *Positive Operator-Valued Measure (POVM)*. An advantage of using POVMs, for example, is that since the POVM elements E_i are positive semidefinite, optimizations over the set of all POVMs can be handled via semidefinite programming techniques.

1.4.3 Evolution of quantum states

We now know how to describe a quantum state $\rho \in \mathcal{D}(\mathcal{X})$, as well how to model a measurement or observation of ρ. The next question we ask is: What kind of operations can we perform on ρ? For example, to a classical bit, we can apply a NOT gate to flip its value. What can we do to a *qu*bit?

In the quantum setting, the set of valid operations on a *closed* (defined shortly) quantum system with state $\rho \in \mathcal{D}(\mathcal{X})$ is the set of unitary operators $U \in \mathcal{U}(\mathcal{X})$. Specifically, U maps ρ to

$$\rho' := U\rho U^\dagger. \tag{1.47}$$

For example, for $\rho \in \mathcal{D}(\mathbb{C}^2)$, i.e. a single qubit, a frequently used set of unitary operators are the *Pauli* operators (where $i := \sqrt{-1} \in \mathbb{C}$)

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.48}$$

Note, for example, that the Pauli X plays the role of a quantum NOT gate, i.e. $X|0\rangle = |1\rangle$ and $X|1\rangle = |0\rangle$.

We said that unitary operations describe the evolution of closed quantum systems above — let us elaborate on this further. A *closed* quantum system is one which does not interact with its environment. Conversely, if a system is not closed, it is called *open*. In this latter case, the set of allowed operations strictly contains $\mathcal{U}(\mathcal{X})$, and is in fact the set of TPCP maps, which we henceforth refer to as *admissible* maps or operations. Despite this, there is a sense in which discussing unitary operations is without loss of generality — this is implied by the Stinespring representation of super-operators and specifically Equation (1.37), which states that any valid TPCP operation on a quantum system A can be simulated by moving to a larger joint system AB, evolving AB via a unitary operator, and subsequently tracing out part of AB. (We discuss joint systems AB further in Section 1.4.4.)

For example, let us consider the process of performing a measurement on A. In order to measure or observe a quantum state in A, one introduces a measurement apparatus, which we think of as system B. To complete the actual measurement, B must interact with A,

implying A is an open system. Thus, if we look at A alone, the action of the measurement on A is not described by a unitary operator, but by a TPCP map. However, if we instead look at AB as a whole, this joint system is now closed, and hence its evolution is described by a unitary operator.

Hamiltonians, and the connection to unitary operations. We said above that the evolution of a (closed) quantum system is described by a unitary operator. Although this is a great abstract description for mathematicians and computer scientists to work with, one should ask the question: Why *unitary* operations? The answer lies, not surprisingly, in physics. Here we define the notion of a *Hamiltonian*, which will play an important role in later chapters such as those involving Hamiltonian complexity.

First, note that any unitary $U \in \mathcal{U}(\mathcal{X})$ can be written as $U = \exp(iH)$ for some $H \in \mathcal{H}(\mathcal{X})$. This is easily seen by taking the spectral decomposition $U = \sum_j e^{i\theta_j}|\psi_j\rangle\langle\psi_j|$, and observing that defining

$$H = \sum_j \theta_j |\psi_j\rangle\langle\psi_j| \tag{1.49}$$

yields $U = e^{iH}$ (see the discussion on operator functions in Section 1.3). The operator H is called a *Hamiltonian*.

Thus, corresponding to each $U \in \mathcal{U}(\mathcal{X})$, there exists an $H \in \mathcal{H}(\mathcal{X})$. Where does H then come from? It turns out that the time evolution of a closed quantum system $|\psi\rangle$ according to H is given by the famous Schrödinger equation,

$$i\hbar \frac{d|\psi\rangle}{dt} = H|\psi\rangle, \tag{1.50}$$

where $\hbar$ denotes Planck's constant (whose value is not of interest here). For a quantum system evolving from time t_1 to t_2, the solution to this equation is given by

$$|\psi(t_2)\rangle = \exp\left(i\frac{t_1 - t_2}{\hbar}H\right)|\psi(t_1)\rangle, \tag{1.51}$$

from which we now see the connection to unitary operators directly.

For this reason, Hamiltonians have been the object of intense study, and there is nowadays an entire field devoted to Hamiltonian complexity (see Section 1.5). The eigenstates $\{|\lambda\rangle\}$ of a Hamiltonian are referred to as its *energy eigenstates*, and the eigenvalue λ corresponding to $|\lambda\rangle$ is the *energy* of state $|\lambda\rangle$. The smallest eigenvalue $\lambda_{\min}$ of H is called the *ground state energy*, and $|\lambda_{\min}\rangle$ the *ground state* of H. Determining the ground state energy of a given H is in general a very difficult problem, as we shall soon see in Section 1.5.

Before closing, we make two final remarks. First, there is another interpretation of the Hamiltonian versus unitary pictures of time evolution presented here which is of interest. The application of any fixed unitary U can be thought of as a *discrete-time* evolution, since by Equation (1.51) it corresponds to evolution by some fixed time t. In the Hamiltonian picture, however, for any fixed Hamiltonian H, one can in principle vary the time of evolution t as desired, resulting in a notion of *continuous-time* evolution.

Finally, in our discussion here we have focused on time-*independent* Hamiltonians. More generally, one can also consider evolution under time-*dependent* Hamiltonians which are allowed to change with time.

1.4.4 Composite quantum systems

Thus far, we have discussed the basics of how to mathematically discuss single quantum systems. Suppose now we have two quantum systems A and B — how do we describe their joint state AB? It turns out that if A and B correspond to complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, then the joint system AB corresponds to the space $\mathcal{X} \otimes \mathcal{Y}$. In other words, if, for example, $\mathcal{X} = \mathcal{Y} = \mathbb{C}^2$, then any $\rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{Y})$ defines a valid two-qubit quantum system.

The simplest examples of two-party systems AB are given by *product states*, which for any given $\rho_A \in \mathcal{D}(\mathcal{X})$ and $\rho_B \in \mathcal{D}(\mathcal{Y})$, are given by $\rho_A \otimes \rho_B$. Such states are uncorrelated between systems A and B. For example, two classical bits in state 00 can be embedded in such a two-qubit quantum state as $|0\rangle \otimes |0\rangle$. For brevity, when discussing pure states, we simply denote this state as $|0\rangle|0\rangle$ or $|00\rangle$. More generally, one can also consider joint states $|\phi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ such as

$$\left|\phi^+\right\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle. \tag{1.52}$$

This state is referred to as a *Bell state*, and possesses a strong degree of *quantum* correlations between systems A and B known as *quantum entanglement*, as discussed further in Section 1.6.

Given a description ρ of the state of a joint system AB, we now require a method for describing the marginal state on A (or B) alone. Specifically, given a composite system $\rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{Y})$, the *reduced state* ρ_A on A (analogously, ρ_B on B) is given by the partial trace operation described in Section 1.3. In other words,

$$\rho_A = \mathrm{Tr}_B(\rho). \tag{1.53}$$

For example, $\mathrm{Tr}_B(\rho_A \otimes \rho_B)$ is simply ρ_A, and $\mathrm{Tr}_B(|\phi^+\rangle\langle\phi^+|) = I/2$. The partial trace is employed here as it is the unique function which correctly produces the measurement statistics for arbitrary observables M measured on A alone.

We close by remarking that our description of two-party composite systems straightforwardly extends to multiple parties: For systems A_1 through A_n corresponding to complex Euclidean spaces $\mathcal{X}_1$ through $\mathcal{X}_n$, the corresponding joint space is given by $\bigotimes_{i=1}^{n} \mathcal{X}_i$.

1.4.5 Quirks of quantum mechanics

Marking a drastic departure from the classical setting, a fundamental result in quantum mechanics is that an unknown quantum state $|\psi\rangle \in \mathcal{X}$ cannot be copied or *cloned*. This is called the *No-Cloning* Theorem [86, 258]. To give a brief flavor of why this holds, we demonstrate a simple proof from Nielsen and Chuang [200] (Box 12.1) for the case regarding the non-existence of a unitary $U \in \mathcal{U}(\mathcal{X} \otimes \mathcal{X})$ achieving the mapping

$$|\psi\rangle_{\mathcal{X}} \otimes |s\rangle_{\mathcal{X}} \mapsto |\psi\rangle_{\mathcal{X}} \otimes |\psi\rangle_{\mathcal{X}}, \tag{1.54}$$

where $|s\rangle$ is some fixed starting state. For sake of contradiction, suppose such a U does exist. Then for vectors $|\psi_1\rangle$,$|\psi_2\rangle$, let

$$|\phi_1\rangle := U(|\psi_1\rangle \otimes |s\rangle) = |\psi_1\rangle \otimes |\psi_1\rangle \tag{1.55}$$
$$|\phi_2\rangle := U(|\psi_2\rangle \otimes |s\rangle) = |\psi_2\rangle \otimes |\psi_2\rangle. \tag{1.56}$$

Then, $\langle\phi_1|\phi_2\rangle = \langle\psi_1|\psi_2\rangle = (\langle\psi_1|\psi_2\rangle)^2$. But the equation $x = x^2$ only has solutions 0 and 1, implying that for general $|\psi_1\rangle$ and $|\psi_2\rangle$, such a U cannot exist. We remark that using the Stinespring representation, this proof is easily adapted to show that even TPCP maps cannot clone non-orthogonal states [261].

1.5 Quantum computational complexity

With the basics of linear algebra and quantum computing under our belts, we can now begin discussing the first central area this thesis studies: Computational complexity theory. This field aims to rigorously classify computational problems based on the inherent difficulty of solving them. Specifically, the central idea here is to ask:

Given a set of resources S, such as a certain amount of space or time in which a computation is to run, what is the class of computational problems which can be solved?

This approach has led to an entire zoo of such *complexity classes* (literally, a zoo [12]), including the ubiquitous classes P and NP. In this section, we review the extension of some of these concepts to the quantum setting. This includes defining the standard quantum circuit model our work is based on, introducing relevant quantum complexity classes, and presenting an exposition of the quantum version of the Cook-Levin theorem [72, 179]. The content of this section is based partly on the excellent surveys of Aharonov and Naveh [22] and Watrous [248], as well as the text of Nielsen and Chuang [200]. We assume background knowledge of basic (classical) computational complexity; the interested reader is referred to the text of Arora and Barak for an introduction [27].

Notation and definitions specific to this section. Throughout our discussion, we encode all computational problems over the binary alphabet $\Sigma := \{0, 1\}$. We say a function $f : \Sigma^* \mapsto \Sigma^*$ is *polynomial-time computable* if there exists a polynomial time deterministic Turing machine which, given any input $x \in \Sigma^*$, outputs $f(x)$. A function $f : \mathbb{N} \mapsto \mathbb{N}$ is called *polynomially-bounded* if there exists a polynomial-time deterministic Turing machine which, on any input $x \in \mathbb{N}$, outputs $1^{f(x)}$. A *language* is a partitioning $\Sigma^* = A_{\text{yes}} \cup A_{\text{no}}$ such that $A_{\text{yes}} \cap A_{\text{no}} = \emptyset$, for $\emptyset$ the empty set. If, more generally, $A_{\text{yes}} \cup A_{\text{no}} \subseteq \Sigma^*$, then we have a *promise problem*. In a promise problem, one assumes the input x satisfies $x \in A_{\text{yes}}$ or $x \in A_{\text{no}}$; if an algorithm solving this promise problem is given input $x \notin A_{\text{yes}} \cup A_{\text{no}}$, we adopt the convention that the algorithm is allowed to err. We remark that promise problems are particularly natural in the quantum setting, as quantum computations are inherently probabilistic processes, and as such, some "margin of error" appears to be needed separating A_{yes} from A_{no}. This is clarified further when introducing our relevant quantum complexity classes.

1.5.1 Quantum circuit model

In Section 1.4.3, we discussed the general types of admissible operations on quantum systems. In the context of complexity theory, however, we require a formal model for specifying and analyzing such operations, for which we employ the standard *quantum circuit model*. To begin, suppose we have a quantum system consisting of n qubits, whose associated complex Euclidean space is $\mathcal{X} = (\mathbb{C}^2)^{\otimes n}$. A *quantum circuit* can be thought of as a directed acyclic graph with n *input* nodes of in-degree zero and out-degree one (i.e. n sources), n *output* nodes of in-degree one and out-degree zero (i.e. n sinks), and a set of "intermediate" nodes or *gates*, each of which has matching in- and out-degree c for some $c \in \Theta(1)$ (where each gate can have a different value of c). Intuitively, the input (output)

nodes are the n input (output) qubits to the circuit, and the intermediate notes are unitary gates acting on $\Theta(1)$ qubits. The edges of the graph correspond to *wires* in the circuit, the direction of which are indicative of the direction of data flow.

For example, three common single-qubit unitary gates mentioned in Section 1.4.3 are the Pauli X, Y, and Z operators, which are specified in the circuit model as:

$$|b\rangle \text{ —}\boxed{X}\text{— } |b \oplus 1\rangle \tag{1.57}$$

$$|b\rangle \text{ —}\boxed{Y}\text{— } (-1)^b i|b \oplus 1\rangle \tag{1.58}$$

$$|b\rangle \text{ —}\boxed{Z}\text{— } (-1)^b|b\rangle \tag{1.59}$$

Here, we assume $b \in \{0, 1\}$; the action of each gate is extended to all single qubit states by linearity. The notation $\oplus$ denotes the XOR operation (i.e. addition modulo 2). On the left of each gate is the input qubit, and on the right is the output qubit.

Two other single-qubit gates, whose importance is discussed shortly, are the *Hadamard* and T (also referred to as $\pi/8$) gates, defined below.

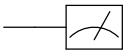
$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \equiv \quad |b\rangle \text{ —}\boxed{H}\text{— } \tfrac{1}{\sqrt{2}}(|0\rangle + (-1)^b|1\rangle) \tag{1.60}$$

$$T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix} \quad \equiv \quad |b\rangle \text{ —}\boxed{T}\text{— } e^{i\frac{\pi \cdot b}{4}}|b\rangle \tag{1.61}$$

A ubiquitous two-qubit gate is the *Controlled-NOT* gate, shown below.

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \equiv \quad \begin{array}{l} |b_1\rangle \text{ —●— } |b_1\rangle \\ |b_2\rangle \text{ —⊕— } |b_2 \oplus b_1\rangle \end{array} \tag{1.62}$$

Finally, a measurement (in the computational basis) in this model is specified by the following.

 (1.63)

Universal gate sets. When it comes to quantifying the *cost* of a circuit, it seems *a priori* that we are in a bind: How do we quantify the cost of an arbitrary gate if there is a continuum of unitary gates to choose from? It would be preferable to have a fixed finite set of gates, each of which is assigned unit cost, and with which we could simulate all other gates. This would yield a rigorous framework in which to quantify the cost of

a circuit. Such a set of unitaries is called a *universal* set, and indeed exists: The set $S := \{H, T, CNOT\}$ is universal. To show this (see, e.g., [200]), one first demonstrates that the $CNOT$ coupled with the set of all one-qubit unitaries is universal in an *exact* sense — any unitary $U \in \mathcal{U}(\mathcal{X})$ can be represented *exactly* using CNOT and single-qubit gates. One then applies the Solovay-Kitaev theorem [173], which yields that for any $U \in \mathcal{U}(\mathbb{C}^2)$ and any $\epsilon > 0$, there exists a $V \in \mathcal{U}(\mathbb{C}^2)$ consisting of the composition of $O(\log^c(1/\epsilon))$ gates from $\{H, T\}$ such that $\| U - V \|_\infty \leq \epsilon$ (here, $c \in \Theta(1)$).

What does such a bound on the spectral norm buy us? Suppose we can substitute the original unitaries $U = U_m \cdots U_1$ in a circuit with unitaries $V = V_m \cdots V_1$ with the promise that $\| U_i - V_i \|_\infty \leq \epsilon$ for all $i \in [m]$ and for ϵ to be chosen as needed. Since we are typically interested in running U on some input $|\psi\rangle$, followed by a measurement according to some POVM, we would like the probability of obtaining any measurement outcome to deviate by at most δ when substituting V for U, where $\delta > 0$ can be chosen as desired. In other words, for all POVM elements M, pure states $|\psi\rangle$, and error parameters $\delta > 0$, we would like that setting ϵ small enough yields that the probability of obtaining outcome M when measuring $U|\psi\rangle$ versus $V|\psi\rangle$ differs by at most δ. Indeed, this is achieved by setting $\epsilon = \delta/(2m)$ and combining the facts that

$$\left|\mathrm{Tr}(MU|\psi\rangle\langle\psi|U^\dagger) - \mathrm{Tr}(MV|\psi\rangle\langle\psi|V^\dagger)\right| \leq 2 \| U - V \|_\infty , \tag{1.64}$$

and

$$\| U_m \cdots U_1 - V_m \cdots V_1 \|_\infty \leq \sum_{j=1}^{m} \| U_j - V_j \|_\infty . \tag{1.65}$$

We refer the reader to [200] for further details.

We close this section by remarking that here we have assumed that quantum circuits are unitary and act on pure state inputs $|\psi\rangle \in \mathcal{X}$; recall from Section 1.4.3 that by the Stinespring representation and Equation (1.37), this is without loss of generality. We refer the reader to the work of Aharonov, Kitaev, and Nisan [21] for a more general model of quantum circuits which directly operates on mixed states, and which explicitly harnesses this connection with the Stinespring representation.

Oracles. A commonly used construct in the setting of quantum circuits is that of an *oracle*. An oracle Q_n (where we more precisely deal with a family of oracles $\{Q_n\}$) can be thought of as a black-box unitary operation encoding some predicate $f : \Sigma^n \mapsto \Sigma$. In the quantum circuit model, this is formalized via the action

$$Q_n|x\rangle|y\rangle = |x\rangle|y \oplus f(x)\rangle, \tag{1.66}$$

for $x \in \Sigma^n$ and $y \in \Sigma$. Each such application of Q_n is called a *query* to the oracle, and we typically think of each query as having unit cost.

Suppose now that we wish to compute some property P of the predicate f; the number of queries to Q_n required to do so is called the *query complexity* of P (relative to Q_n). Perhaps the most well-known example of this in the quantum setting is Grover's algorithm [113], which shows how to compute the OR function $\bigvee_{i=1}^{2^n} f(i)$ with high probability using $O(\sqrt{2^n})$ queries to Q_n, a quadratic improvement over the classical setting. Although the query model may *a priori* seem restricted, the model is nevertheless important; Shor conceived his factoring algorithm [224], for example, by studying Simon's algorithm [225, 226] from the quantum query model.

1.5.2 Standard quantum complexity classes: BQP and QMA

Recall that in complexity theory, we classify computational problems into *complexity classes* depending on the resources capable of solving them. The classes P and NP are two such classes, forming two cornerstones of classical complexity theory. We now discuss the natural quantum analogues of these classes, BQP and QMA. (More precisely, BQP and QMA are generalizations of BPP and MA.) For completeness, we recall the definitions of P and NP below.

Definition 1.1 (P)**.** *A promise problem $A = (A_{\text{yes}}, A_{\text{no}})$ is in P if and only if there exists a deterministic polynomial-time Turing machine M which on input $x \in A_{\text{yes}}$, accepts, and on input $x \in A_{\text{no}}$, rejects.*

Definition 1.2 (NP)**.** *A promise problem $A = (A_{\text{yes}}, A_{\text{no}})$ is in NP if and only if there exists a deterministic polynomial-time Turing machine M and a polynomial p, such that on input $x \in \Sigma^*$:*

- *If $x \in A_{\text{yes}}$, then there exists a proof $y \in \Sigma^{p(|x|)}$ such that M accepts (x, y).*
- *If $x \in A_{\text{no}}$, then for all proofs $y \in \Sigma^{p(|x|)}$, M rejects (x, y).*

Now, since we have defined our complexity theoretic model for quantum computing based on the quantum circuit model, we next require the notion of a polynomial-time uniform family of quantum circuits. Specifically, since the length of input $x \in \Sigma^*$ to a computational problem is allowed to vary, whereas the input size to a given circuit is fixed, we require a method for "scaling" our circuits up to match the length of arbitrary input $x \in \Sigma^*$.

Definition 1.3 (Polynomial-time uniform family of quantum circuits)**.** *A set of quantum circuits $\{Q_n\}$ is polynomial-time uniform if there exists a polynomial-time deterministic Turing machine, which on input 1^n, outputs a description of Q_n.*

We now define BQP [46], which stands for *Bounded-Error Quantum Polynomial Time*, and which is intuitively the set of promise problems which can be efficiently solved with high probability on a quantum computer. For both BQP and QMA, we henceforth say a quantum circuit Q *accepts* input x (where x can be either a classical string or quantum state) if running Q on input x and subsequently measuring a designated output qubit of Q in the computational basis yields outcome 1.

Definition 1.4 (BQP)**.** *A promise problem $A = (A_{\text{yes}}, A_{\text{no}})$ is in BQP if and only if there exists a polynomial q and a polynomial-time uniform family of quantum circuits $\{Q_n\}$, where Q_n takes as input a string $x \in \Sigma^*$ with $|x| = n$, and $q(n)$ ancilla qubits in state $|0\rangle^{\otimes q(n)}$, such that:*

- *(Completeness) If $x \in A_{\text{yes}}$, then Q_n accepts input x with probability at least $2/3$.*
- *(Soundness) If $x \in A_{\text{no}}$, then Q_n accepts input x with probability at most $1/3$.*

Note that if we replace the uniform quantum circuit family above with a uniform *classical* circuit family which takes as input both x and a polynomial-size string y chosen uniformly at random, then we are reduced to BPP. Like BPP, the completeness and soundness parameters $2/3$ and $1/3$ above can straightforwardly be amplified to values exponentially close to 1 and 0 simply by running the verification procedure Q independently polynomially many times in parallel, accepting if and only if the majority of runs accepted, and applying the Chernoff bound. We remark that BPP $\subseteq$ BQP follows since probabilistic classical computations can be simulated with quantum circuits (see, e.g. [248]). The decision versions of the factoring and discrete logarithm problems are, for example, not known to be in BPP, but are in BQP due to Shor's algorithm [224].

We next define QMA, or *Quantum Merlin Arthur*, a quantum generalization of NP.

Definition 1.5 (QMA)**.** *A promise problem $A = (A_{\text{yes}}, A_{\text{no}})$ is in QMA if and only if there exist polynomials p, q and a polynomial-time uniform family of quantum circuits $\{Q_n\}$, where Q_n takes as input a string $x \in \Sigma^*$ with $|x| = n$, a quantum proof $|y\rangle \in (\mathbb{C}^2)^{\otimes p(n)}$, and $q(n)$ ancilla qubits in state $|0\rangle^{\otimes q(n)}$, such that:*

- *(Completeness) If $x \in A_{\text{yes}}$, then there exists a proof $|y\rangle \in (\mathbb{C}^2)^{\otimes p(n)}$ such that Q_n accepts $(x, |y\rangle)$ with probability at least $2/3$.*

- *(Soundness) If $x \in A_{\text{no}}$, then for all proofs $|y\rangle \in (\mathbb{C}^2)^{\otimes p(n)}$, Q_n accepts $(x, |y\rangle)$ with probability at most $1/3$.*

It is often helpful to think of $|y\rangle$ above as a proof sent by an all-powerful but untrustworthy prover Merlin, who claims $x \in A_{\text{yes}}$, and to correspondingly interpret $\{Q_n\}$ as an honest but computationally bounded verifier Arthur, whose job it is to verify the correctness of Merlin's proof. We are not overly fond of the names Merlin and Arthur, and as such, prefer to simply refer to both parties in this interpretation as being the prover and verifier, respectively. As an aside, we remark that QMA was originally known as Bounded Error Quantum NP (BQNP) [171].

Note now that if we instead ask in the definition of QMA that $y \in \Sigma^{p(|x|)}$, then the corresponding complexity class is known as quantum-classical Merlin-Arthur (QCMA) [22, 156, 6, 11, 39, 18, 257]. (QCMA is also known by the name Merlin-Quantum-Arthur (MQA), as suggested by Watrous [248].) Finally, if y is classical *and* we replace $\{Q_n\}$ with a classical circuit family of the type used in defining BPP, then the class we obtain is Merlin-Arthur (MA) [33].

Error reduction for QMA. Like BQP, the completeness and soundness parameters in the definition of QMA can be amplified to values exponentially close to 1 and 0, respectively. However, the arguments employed here are not as straightforward as in the case of BQP. For QMA, there are two approaches for achieving error reduction, which we refer to as *weak* and *strong* error reduction, and which we now discuss.

Weak or *standard* error reduction runs analogously to the case of BQP, i.e. by running the verification protocol some number of times m in parallel and taking a majority vote. However, since from Section 1.4.5, we know that unknown quantum states cannot be cloned, the verifier must ask the prover for multiple copies of the proof $|y\rangle$, one for each of the m parallel runs of the protocol. If the verifier is honest, the proof sent for the new protocol is a product state $|y'\rangle = |y\rangle^{\otimes m} \in (\mathbb{C}^2)^{\otimes p(|x|)\cdot m}$, in which case the m runs of the verification protocol are independently and identically distributed Bernoulli trials, and the Chernoff bound can be applied. However, if we have a NO-instance, i.e. $x \in A_{\text{no}}$, then in a desperate attempt to trick the verifier into thinking $x \in A_{\text{yes}}$, the prover may elect to cheat by sending a proof $|y'\rangle$ which deviates from this product state structure. Can we still apply the Chernoff bound argument here?

It turns out the answer is *yes*, the intuition for which we now sketch. (A detailed proof can be found in [22].) Specifically, let V denote the original verification protocol. Then, given any $|y'\rangle \in (\mathbb{C}^2)^{\otimes p(|x|)\cdot m}$, we adopt the following view: On the first $p(|x|)$ proof qubits,

we run the first copy of V, measure and read the output qubit, and subsequently discard these $p(|x|)$ qubits. Note that the reduced state of $|y'\rangle$ on these first $p(|x|)$ qubits before running V is simply a convex mixture of proofs $|y\rangle \in (\mathbb{C}^2)^{\otimes p(|x|)}$; thus, by the soundness property of the QMA protocol, the probability of acceptance in this first run is at most $1/3$. We can iterate this argument over each of the remaining $m-1$ copies of V, each time obtaining a probability of accepting of at most $1/3$. It follows that a majority vote, coupled with the Chernoff bound, yields the desired error reduction.

Finally, although weak error reduction is simple, its disadvantage is that it requires an increase in the proof size, since the prover must send multiple copies of the original proof. Is it possible to reduce the error *without* increasing the proof length? Remarkably, Marriot and Watrous have shown [191] that the answer is *yes*. The rough idea here is best illustrated in the case of a zero-error verifier V, i.e. where the completeness and soundness parameters are 1 and 0, respectively. Specifically, let V be a zero-error verifier V, and $|y\rangle$ the prover's proof for some instance $x \in A_{\text{yes}}$. Then, if we run V on $(x, |y\rangle)$ and measure the output qubit, we will see outcome 1 with certainty. Thus, the measurement does not alter the output state of V. Further, if we now run V in reverse and measure the ancillary qubits of V, they should read all zeroes with certainty, implying this second measurement also does not alter the state being measured. In fact, we can repeat this back and forth process as many times as we like, each time obtaining the same "good" measurement outcomes.

What happens now if we do not have a zero-error QMA verifier V, and have a NO instance $x \in A_{\text{no}}$? In this case, the output qubit of $V|x\rangle \otimes |y\rangle$ must yield outcome 1 with probability at most $1/3$ — in other words, measuring this qubit now disturbs the state $V|x\rangle \otimes |y\rangle$. Moreover, when we next apply $V^\dagger$ and measure the ancilla qubits, since V is unitary, the outcome cannot be the all-zeroes string with non-negligible probability, again disturbing the state. Intuitively, by repeating this back-and-forth procedure, we thus quickly amplify the likelihood of obtaining "bad" measurement outcomes in this process. In our opinion, the entire process can be thought of as analogous to a spinning top — if the top wobbles badly enough to begin with (if $x \in A_{\text{no}}$), the spinning motion (the back and forth measurement process) quickly sends the top out of control.

1.5.3 BQP and QMA in further depth

As QMA plays an important role in this thesis, we now further discuss its properties, variants, and complete problems. Along the way, we also mention some further properties of BQP.

First, we have

$$\text{NP} \subseteq \text{MA} \subseteq \text{QCMA} \subseteq \text{QMA} \subseteq \text{PP}. \tag{1.67}$$

Here, PP is defined analogously to BPP, except that when input $x \in A_{\text{yes}}$, then the verifier accepts with probability strictly larger then $1/2$; if $x \in A_{\text{no}}$, the verifier accepts with probability at most $1/2$. The second of the containments above follows since a QCMA verifier can choose to act classically. The third containment holds since a QMA verifier can force a given quantum proof to encode a classical string by preceding the verification procedure with a measurement in the computational basis. Finally, the fourth containment has an elegant proof via the strong error reduction technique of Marriott and Watrous [191], and was originally proven by Kitaev and Watrous [172].

Regarding BQP, we have that

$$\text{BPP} \subseteq \text{BQP} \subseteq \text{QMA}, \tag{1.68}$$

where the second containment follows since the verifier can simple flush the prover's proof down the toilet and run the BQP circuit instead. Combining Equations (1.67) and (1.68) yields BQP $\subseteq$ PP; we remark that this containment was directly proven by Adleman, DeMarrais, and Huang [14] and Fortnow and Rogers [100]. Marriott and Watrous have shown that BQP $=$ $\text{QMA}_{\log}$ [191], where $\text{QMA}_{\log}$ is QMA with a logarithmic size proof. The classical version of this equality might be written P $=$ $\text{NP}_{\log}$, i.e. NP with logarithmic size proofs is contained in P. Finally, it is well-known that in the classical setting, BPP $\subseteq$ Σ_2^p [227, 177], for Σ_2^p the second level of the polynomial hierarchy PH. Whether BQP $\subseteq$ PH, however, remains a major open question [8, 93, 9].

One-sided error. Next, we discuss the one-sided error versions of MA, QCMA, and QMA. Specifically, let MA_1, QCMA_1, and QMA_1 be defined as MA, QCMA, and QMA, respectively, except with completeness 1 in each case. In other words, if $x \in A_{\text{yes}}$, the verifier for the new classes accepts with certainty. Zachos and Furer have shown that MA $=$ MA_1 [262] (see also Goldreich and Zuckerman [116]), and more recently, Jordan, Kobayashi, Nagaj, and Nishimura have proven that QCMA_1 $=$ QCMA [159]. Whether QMA_1 $=$ QMA, however, remains an interesting open question, particularly since both QCMA and QIP(3) in the chain QCMA $\subseteq$ QMA $\subseteq$ QIP(3) allow one-sided error [172]. Here, QIP(k) is the class of promise problems having *Quantum Interactive Proofs* with k rounds, meaning it is a generalized version of QMA in which k quantum messages are passed back and forth between prover and verifier. For example, BQP $=$ QIP(0), QMA $=$ QIP(1), and QIP(3) consists of a message from prover to verifier, followed by a message from verifier to prover, and a final message back from the prover to the verifier. Aaronson has demonstrated a quantum oracle relative to which QCMA_1 $\subset$ QCMA and QMA_1 $\subset$ QMA [7].

Complete problems. We now move to arguably one of the most important questions for any complexity class: *What problems characterize, or are complete for QMA?* In general, the set of QMA-complete problems is not yet nearly as rich as that for its classical cousin, NP. The historically first QMA-complete problem was the local Hamiltonian problem (first presented by Kitaev at [170], and later written up in [171]), which is a natural generalization of the NP-complete problem of classical constraint satisfaction, and relevant from a physics perspective. In fact, we devote Section 1.5.4 entirely to this problem and its variants, and thus do not discuss it further here.

Perhaps the second-most studied and natural QMA-complete problem is the Consistency problem for local density matrices of Liu [182]. In this problem, one is given a classical description of a set of density matrices ρ_S, each acting on a subset $S \subseteq [n]$ qubits for $|S| = k$ and $k \in \Theta(1)$. The question is whether there exists a globally consistent n-qubit state ρ such that $\mathrm{Tr}_{[n]\setminus S}(\rho) = \rho_S$ for all S. The proof of QMA-hardness for $k = 2$ follows via a polynomial-time Turing or Cook reduction involving convex programming from the 2-local Hamiltonian problem [182]; the reduction in the reverse direction was later given by Liu in [183], and goes via a strong theorem of alternatives in semidefinite programming. Other physically motivated variants of the Consistency problem have also been shown to be QMA-complete: The variant involving fermions, known as the N-representability problem, was shown QMA-complete by Liu, Christandl, and Verstraete [184], as well as its bosonic counterpart by Wei, Mosca, and Nayak [253].

What other QMA-complete problems are known? Given a classical description of a quantum circuit, the problem of determining whether it is "close" to the identity, known as the Identity Check problem, was shown QMA-complete by Janzing, Wocjan, and Beth [157]. Rosgen [215] has shown that a similar problem where one is asked whether a given quantum circuit is close to a linear isometry is QMA-complete. Finally, Beigi and Shor [40] have proposed a QMA-complete quantum generalization of the Clique problem, which asks: Given an (entanglement-breaking) channel Φ, do there exist k quantum states ρ which are distinguishable without error after passing through the channel?

Multiple provers. QMA is a proof system with a single prover and verifier. A curiosity emerges when we ask the question: What happens to the power of the proof system if we introduce a *second* prover? In other words, what if there are two provers, P_1 and P_2, who send a joint proof of the form $|\psi_P\rangle \otimes |\psi_Q\rangle$ to the verifier? Interestingly, unlike the classical setting where having two provers is trivially equivalent to having a single prover, in the quantum setting, the possibility of *entanglement* between the two proofs (entanglement is introduced in Section 1.6) makes this a non-trivial question. This class

is called QMA(2) [175]. Why should it be of any interest? Perhaps surprisingly, Blier and Tapp [48] have shown that all languages in NP have very short proofs in this model; specifically, it suffices for P_1 and P_2 to send proofs $|\psi_{P_1}\rangle$ and $|\psi_{P_2}\rangle$, respectively, consisting of just $O(\log n)$ qubits each. The reader is referred to Chapter 4 for formal definitions and details regarding this model, where it is studied in further depth.

1.5.4 Local Hamiltonian complexity: An overview

In Section 1.5.3, we initiated our discussion of QMA-complete problems, and stated that the first known such problem was the local Hamiltonian problem. As this problem features heavily in Chapters 2 and 3, we now discuss it in further depth. We begin by defining the problem, and follow by demonstrating how it generalizes the canonical NP-complete problem MAX-SAT. We then discuss some of its variants and its history with respect to the field of complexity theory. Later in Section 1.5.5, we give Kitaev's proof that the 5-local Hamiltonian problem is QMA-complete.

Beginning with definitions, the local Hamiltonian problem (LH) was introduced by Alexei Kitaev [170, 171], and can intuitively be thought of as follows: Given a "succint" representation of a "large" Hamiltonian H, what is H's smallest eigenvalue? Of course, the obvious approach to answering this question is to diagonalize H — however, the catch is that while H is a $2^n \times 2^n$-dimensional matrix, the succinct encoding we are given of H consists of $\text{poly}(n)$ bits. In other words, a simple diagonalization approach would take time exponential in the input size.

Let us now define LH more formally. To do so, we first define the term *k-local Hamiltonian*.

Definition 1.6. *An operator $H \in \mathcal{H}(\mathcal{B}^{\otimes n})$ is called a k-local Hamiltonian if it can be written*

$$H = \sum_{j=1}^{r} H_j, \tag{1.69}$$

where $\{H_j\}_{j=1}^{r} \subseteq \mathcal{H}(\mathcal{B}^{\otimes k})$ is a collection of local Hamiltonian terms, such that each H_j acts non-trivially on some subset $S_j \subseteq [n]$ of at most k qubits and satisfies $0 \preceq H_j \preceq I$. Note: In Equation (1.69), we adopt the convention that each H_j acts as the identity on all qubits in the set $[n] \backslash S_j$.

Note that although we define H as acting on qubits above, the definition extends straightforwardly to the case of higher-dimensional local systems. Intuitively, the definition above

says that a k-local Hamiltonian H can be expressed as a sum of "smaller" Hermitian operators H_j, each of which is restricted to act non-trivially on at most k out of n qubits.

We now phrase the k-LH problem. We remark that later, in Chapter 2, we shall formulate k-LH in a slightly different manner; the definition below is, however, arguably more natural and thus better suited to an introductory section.

Problem 1.7 (k-Local Hamiltonian (k-LH) [171])**.** *Given as input:*

1. *A k-local Hamiltonian H acting on n qubits, specified as a collection of local Hamiltonian terms $\{H_j\}_{j=1}^{r} \subseteq \mathcal{H}(\mathcal{B}^{\otimes k})$ (i.e. as a collection of $(2^k \times 2^k)$-dimensional matrices H_j) where $k \in \Theta(1)$,*
2. *Threshold parameters $a, b \in \mathbb{R}$, such that $0 \leq a < b$ and $(b-a) \geq 1$,*

decide, with respect to the complexity measure $\langle H \rangle + \langle a \rangle + \langle b \rangle$:

1. *If $\lambda_{\min}(H) \leq a$, output YES.*
2. *If $\lambda_{\min}(H) \geq b$, output NO.*

Note that often k-LH is phrased with $(b-a) \geq 1/p(n)$ for some polynomial p; such an inverse polynomial gap can straightforwardly be boosted to the constant 1 above by defining H to have $p(n)$ many copies of each local term H_j [248].

Although it may not be *a priori* obvious, k-LH generalizes the canonical NP-complete problem MAX-k-CSP, where CSP stands for *Constraint Satisfaction Problem* (of which a special case is the more familiar problem MAX-k-SAT). To see this, recall that in MAX-k-CSP, one is given a set of Boolean functions, $c_i : \{0,1\}^k \mapsto \{0,1\}$ (note the c_i are not restricted to be of any particular form such as conjunctive normal form), where each c_i acts on k out of n possible bits. We then ask: What is the largest number of clauses c_i we can satisfy with a Boolean assignment to the n bits? To embed this problem into k-LH, we design a k-local Hamiltonian H acting on n qubits as follows. For each clause c_i, define a $2^k \times 2^k$-dimensional diagonal matrix $H_{c_i} \in \mathcal{H}(\mathcal{B}^{\otimes k})$ such that $H_{c_i}(m,m) = 0$ if the binary representation of m is a satisfying assignment for clause c_i; otherwise, $H_{c_i}(m,m) = 1$. In other words, for $x \in \{0,1\}^k$, $\mathrm{Tr}(H_{c_i}|x\rangle\langle x|) = 0$ if x satisfies c_i, and $\mathrm{Tr}(H_{c_i}|x\rangle\langle x|) = 1$ otherwise, i.e. failing assignments are given an *energy penalty*. To now see that the optimal value of our MAX-k-CSP instance corresponds to the smallest eigenvalue of $H = \sum_c H_{c_i}$, we use the fact that since all the H_{c_i} are diagonal, they commute and thus simultaneously diagonalize. Hence, H has integer eigenvalues. Moreover, since the H_{c_i} are simultaneously diagonal in the computational basis, the smallest eigenvalue of H equals the minimum number of unsatisfied clauses over all n-qubit computational basis states. It follows that k-LH generalizes MAX-k-CSP, and thus k-LH is NP-hard. This raises the natural question: *Could* k-LH *be a canonical QMA-complete quantum constraint satisfaction problem?*

Variants of k-LH **and a brief history.** It turns out that k-LH is indeed QMA-complete; Kitaev [170, 171] showed the problem to be in QMA for $k \geq 1$ and QMA-hard for $k \geq 5$. The proof of QMA-hardness was inspired by earlier ideas of Feynman [171, 97], and can be thought of as exploiting Feynman's ideas to adapt the classical Cook-Levin theorem in a non-trivial fashion to the quantum setting. The fact that 3-LH is also QMA-complete was shown subsequently by Kempe and Regev [164] (an alternate proof was later also given by Nagaj and Mozes [199]). Finally, Kempe, Kitaev, and Regev showed [163] that even 2-LH is QMA-complete. Note that 1-LH is in P, since one can simply optimize for each 1-local term independently. Although these results are interesting from a complexity theoretic perspective, a more natural question from a physics perspective is whether such QMA-hardness results can be shown even if the QMA-hard classes of local Hamiltonians arising in the reductions employed correspond to *physical* quantum systems in nature [202, 20, 198, 223]. Along these lines, Oliveira and Terhal next showed [202] that 2-LH with the Hamiltonians restricted to nearest-neighbor interactions on a 2D grid is still QMA-complete. Furthermore, in stark contrast to the classical case of MAX-2-CSP on the line (which is in P), Aharanov, Gottesman, Irani and Kempe [20] showed that 2-LH with nearest-neighbor interactions on the line is also QMA-complete if the local systems have dimension at least 12 (Nagaj later improved this to 11 states per particle [198]).

Although this thesis focuses on the general local Hamiltonian problem as defined in Definition 1.7, for completeness, we now mention a few interesting variants of LH which have also been studied. First, Bravyi and Vyalyi showed that the variant of 2-LH (with local systems of arbitrary, but constant, dimension) in which all local Hamiltonian terms H_j pairwise commute is in NP. This result was extended to the case of 3-LH on qubits by Aharonov and Eldar [19]. Bravyi [55] introduced a variant of k-LH known as Quantum k-SAT, in which each local Hamiltonian term H_j is a projector, and in which the threshold a is set to 0. We remark that in the YES case of such a setup, the local Hamiltonian is referred to as *frustration-free*, since the optimal assignment lies in the null space of every interaction term. Bravyi then showed that, like classical 2-SAT, Quantum 2-SAT is in P (whereas recall 2-LH is QMA-complete) [55]. In contrast, Quantum 4-SAT is QMA_1-complete (recall QMA_1 is the one-sided error analog of QMA) [55]. Whether Quantum 3-SAT on qubits is QMA_1-complete remains an intriguing open question (see Reference [199]). Next, there has been a line of work on so-called *stoquastic* local Hamiltonians [56, 58, 59, 183, 158]. Specifically, the Stoquastic k-SAT problem, defined the same as Quantum k-SAT except that all local projectors have real non-negative matrix elements when expressed in the computational basis, was shown to be in MA for $k \geq 1$, and MA-complete for $k \geq 6$ [56, 59]. (Incidentally, this was the first non-trivial example of an MA-complete promise problem.) The problem Stoquastic LH-MIN, defined as k-LH except where each local Hamiltonian

constraint H_j has real non-positive off-diagonal matrix elements in the computational basis, was shown complete for the class StoqMA [56] for $k \geq 2$. Here, StoqMA is a variant of QMA in which the verifier is restricted to preparing qubits in the states $|0\rangle$ and $|+\rangle$, performing classical reversible gates, and measuring in the Hadamard (i.e. $|+\rangle, |-\rangle$) basis. Note that MA $\subseteq$ StoqMA $\subseteq$ QMA. Finally, variations of LH with symmetry constraints have been studied from a complexity theoretic perspective in, for example, [117, 161].

Connection to physics. Although we have primarily discussed LH from a complexity theoretic viewpoint involving quantum constraint satisfaction, the initial motivation for studying LH comes of course from physics. Indeed, the study of the local Hamiltonian problem is part of the more general field of Hamiltonian Complexity, whose aim is to understand how difficult it is to simulate physical systems. In particular, LH can be phrased as a special case of the more general Simulation Problem [204], which roughly asks the following: Given a description of a Hamiltonian H, an initial state ρ, an observable M, and a time $t \in \mathbb{C}$, estimate the expectation

$$\mathrm{Tr}\left[M \frac{(e^{iHt})^\dagger \rho e^{iHt}}{\mathrm{Tr}\left((e^{iHt})^\dagger \rho e^{iHt}\right)}\right]. \tag{1.70}$$

The local Hamiltonian problem is recovered by choosing H as a local Hamiltonian, setting $M = H$, $\rho = I/\mathrm{Tr}(I)$, and considering $t = i\beta$ for $\beta \in \mathbb{R}$ and $\beta \to \infty$. We refer the reader to the survey of Osborne for further details [204].

1.5.5 Kitaev's quantum Cook-Levin theorem

In Section 1.5.4, we discussed the local Hamiltonian problem (LH) and its variants. As Chapter 3 heavily exploits the structure and details of Kitaev's quantum version of the Cook-Levin theorem, i.e. his proof that 5-LH is QMA-complete, we present the latter here. This requires two steps: One first shows that k-LH $\in$ QMA for $k \geq 1$. One then shows that k-LH is QMA-hard for $k \geq 5$. Our discussion is based on a project completed by the present author for a graduate course on quantum complexity theory at the University of Waterloo [104], and follows the text of Kitaev, Shen, and Vyalyi [171] closely. The reader is referred to the survey of Aharonov and Naveh for an alternate exposition [22].

Local Hamiltonian is in QMA

We begin by showing that k-LH $\in$ QMA for any constant k. Specifically, for any YES-instance (H, a, b) of k-LH with k-local Hamiltonian $H = \sum_{j=1}^{r} H_j \in \mathcal{L}(\mathcal{B}^{\otimes n})$, we show that

there exists a poly-size quantum proof $|\psi\rangle$ and a poly-size quantum verification circuit V, such that a single-qubit measurement on $V|\psi\rangle$ yields 1 with high probability.

First, the quantum proof is constructed as $|\psi\rangle \in \mathbb{C}^r \otimes \mathcal{B}^{\otimes n} \otimes \mathcal{B}$ as:

$$|\psi\rangle = \left(\frac{1}{\sqrt{r}} \sum_{j=1}^{r} |j\rangle\right) \otimes |\eta\rangle \otimes |0\rangle, \tag{1.71}$$

for $\{|j\rangle\}_{j=1}^r$ an orthonormal basis for $\mathbb{C}^r$, and $|\eta\rangle$ an eigenvector corresponding to some eigenvalue λ of H. We call the first register of $|\psi\rangle$ the *index* register, the second the *proof* register, and the last the *answer* register.

To define the verification procedure V, recall that $H = \sum_{j=1}^r H_j$, where each H_j acts on the set of qubits denoted by S_j. Suppose H_j has spectral decomposition $H_j = \sum_s \lambda_s |\lambda_s\rangle\langle\lambda_s|$. Then, define unitary W_j acting on the proof and answer registers, i.e. $W_j \in \mathcal{U}(\mathcal{B}^n \otimes \mathcal{B})$, such that

$$W_j \left(|\lambda_s\rangle \otimes |0\rangle\right) = |\lambda_s\rangle \otimes \left(\sqrt{\lambda_s}|0\rangle + \sqrt{1-\lambda_s}|1\rangle\right). \tag{1.72}$$

Observe that one can implement this operation as follows. First, run phase estimation on $\exp(iH_j)$ to extract λ_s to some ancilla register. Despite the fact that simulating $\exp(iH_j)$ can in general be costly, in our case, since $|S_j|$ is constant, the simulation can be done efficiently. Conditioned on the value of the ancilla, we then rotate the answer register to obtain the desired superposition, and finally uncompute λ_s in the ancilla. Define now unitary $V := \sum_{j=1}^r |j\rangle\langle j| \otimes W_j$.

Having defined $|\psi\rangle$ and V, the verification procedure now proceeds as follows:

1. Apply V to $|\psi\rangle$.
2. Measure the answer register and return the result.

Let us analyze the probability of measuring 1 in the answer register with this procedure. If we assume the index register is implicitly measured at the end of the verification, then we can think of Step 1 above as using the index register to choose an index j uniformly at random, followed by applying W_j to the proof register. Then, we can analyze the probability that this procedure returns 1 as follows:

$$\Pr(\text{output } 1) = \sum_{j=1}^{r} \frac{1}{r} \Pr(\text{output } 1 \mid W_j \text{ is applied}), \tag{1.73}$$

where one has

$$\begin{aligned}\Pr(\text{output } 1 \mid W_j \text{ is applied}) &= \operatorname{Tr}\left[(I_{\mathcal{B}^{\otimes n}} \otimes |1\rangle\langle 1|)W_j(|\eta\rangle\langle\eta| \otimes |0\rangle\langle 0|)W_j^\dagger\right] \quad (1.74)\\ &= (\langle\eta| \otimes \langle 0|)W_j^\dagger(I_{\mathcal{B}^{\otimes n}} \otimes |1\rangle\langle 1|)W_j(|\eta\rangle \otimes |0\rangle). \quad (1.75)\end{aligned}$$

The projector $|1\rangle\langle 1|$ above acts on the answer register. To simplify this, rewrite $|\eta\rangle$ in the eigenbasis of H_j, i.e. $|\eta\rangle = \sum_s \alpha_s|\lambda_s\rangle$, and observe that

$$\begin{aligned}(I_{\mathcal{B}^{\otimes n}} \otimes \langle 1|)W_j(|\eta\rangle \otimes |0\rangle) &= (I_{\mathcal{B}^{\otimes n}} \otimes \langle 1|)W_j\left(\sum_s \alpha_s|\lambda_s\rangle \otimes |0\rangle\right) \quad (1.76)\\ &= (I_{\mathcal{B}^{\otimes n}} \otimes \langle 1|)\left[\sum_s \alpha_s|\lambda_s\rangle \otimes \left(\sqrt{\lambda_s}|0\rangle + \sqrt{1-\lambda_s}|1\rangle\right)\right]\\ &= \sum_s \alpha_s\left(\sqrt{1-\lambda_s}\right)|\lambda_s\rangle. \quad (1.77)\end{aligned}$$

Substituting this into Equation (1.75), we obtain:

$$\begin{aligned}\Pr(\text{output } 1 \mid W_j) &= \left(\sum_t \alpha_t^*\left(\sqrt{1-\lambda_t}\right)\langle\lambda_t|\right)\left(\sum_s \alpha_s\left(\sqrt{1-\lambda_s}\right)|\lambda_s\rangle\right)\\ &= \sum_s (1-\lambda_s)\,|\alpha_s|^2 \quad (1.78)\\ &= 1 - \sum_s \lambda_s\,|\alpha_s|^2 \quad (1.79)\\ &= 1 - \langle\eta|H_j|\eta\rangle, \quad (1.80)\end{aligned}$$

where we have used the fact that $\sum_s |\alpha_s|^2 = 1$. Substituting this into Equation (1.73) finally yields:

$$\Pr(\text{output } 1) = \sum_{j=1}^r \frac{1}{r}1 - \langle\eta|H_j|\eta\rangle = 1 - \frac{1}{r}\langle\eta|\left(\sum_{j=1}^r H_j\right)|\eta\rangle = 1 - \frac{1}{r}\langle\eta|H|\eta\rangle. \quad (1.81)$$

Recalling that we chose η to be an eigenvector of H with some eigenvalue λ, we have that if H corresponds to a YES instance (i.e. there exists $\lambda \leq a$), it follows that we can choose η such that our verification procedure returns 1 with probability $1 - r^{-1}\lambda \geq 1 - r^{-1}a$. On the other hand, if H corresponds to a NO instance (i.e. for all λ, we have $\lambda \geq b$), we have $\Pr(\text{output } 1) \leq 1 - r^{-1}b$. Since the probabilities in the YES and NO cases differ by an inverse polynomial in the input size, we can apply the error reduction techniques for QMA discussed in Section 1.5.2 to conclude that LH $\in$ QMA.

5-local Hamiltonian is hard for QMA

We next show that 5-local Hamiltonian is QMA-hard. To do so, we show a polynomial-time many-one or Karp reduction from an arbitrary problem in QMA to 5-LH.

To begin, let P be a promise problem in QMA, and let $V = V_L V_{L-1} \dots V_1$ be a verification circuit for P composed of unitaries V_k. Without loss of generality, we assume each V_k acts on pairs of qubits. We assume $V \in \mathcal{U}(\mathcal{B}^{\otimes m} \otimes \mathcal{B}^{\otimes N-m})$, where the m-qubit register contains the proof V verifies, and the remaining qubits are ancilla qubits.

Our goal is to define a 5-local Hamiltonian H that will have a small eigenvalue if and only if there exists a proof $|\psi\rangle \in \mathcal{B}^{\otimes m}$ causing V to accept with high probability. Kitaev's idea [171] was to exploit the structure of V by forcing the minimizing eigenvector of H to "simulate" the action of V. To do so, let H act on $\mathcal{B}^{\otimes m} \otimes \mathcal{B}^{\otimes N-m} \otimes \mathbb{C}^{L+1}$, which is simply the initial space V acts on, tensored with an $(L+1)$-dimensional *counter* or *clock* register. This clock register will "keep track of time" in the simulation, i.e. a value of k in the register will correspond to having "applied" $V_1 \dots V_k$. For clarity of exposition, where necessary, we label the three registers H acts on as p for proof, a for ancilla, and c for clock, respectively.

Having defined the space H acts on, we now define H itself:

$$H := H_{\text{in}} + H_{\text{prop}} + H_{\text{out}}, \tag{1.82}$$

with the terms H_{in}, H_{prop}, and H_{out} defined as follows (intuitive explanations to follow). Let

$$H_{\text{in}} := I_p \otimes (I_a - |0 \dots 0\rangle\langle 0 \dots 0|_a) \otimes |0\rangle\langle 0|_c. \tag{1.83}$$

Note that the projector $(I_a - |0 \dots 0\rangle\langle 0 \dots 0|_a)$ is used here for simplicity of exposition; the same analysis holds if we instead use the 1-local constraint $\sum_{i=1}^{N-m} (|1\rangle\langle 1|_i)_c$ (where the ith projector acts on the ith ancilla qubit) — hence, we do not violate the constraint that H be 5-local. Next, H_{out} is defined as

$$H_{\text{out}} := (|0\rangle\langle 0| \otimes I_{\mathcal{B}^{\otimes m-1}})_p \otimes I_a \otimes |L\rangle\langle L|_c. \tag{1.84}$$

Finally, define H_{prop} as

$$\begin{aligned} H_{\text{prop}} &:= \sum_{j=1}^{L} H_j, \quad \text{where} \\ H_j &:= -\frac{1}{2} V_j \otimes |j\rangle\langle j-1|_c - \frac{1}{2} V_j^\dagger \otimes |j-1\rangle\langle j|_c + \qquad (1.85) \\ &\quad \frac{1}{2} I \otimes (|j\rangle\langle j| + |j-1\rangle\langle j-1|)_c. \qquad (1.86) \end{aligned}$$

Each of the terms H_{in}, H_{out}, and H_{prop} allow us to "force" the minimizing eigenvector of H to "simulate" V as follows. Recall that our goal is to have $\langle\eta|H|\eta\rangle$ for some $|\eta\rangle \in \mathcal{B}^{\otimes m} \otimes \mathcal{B}^{\otimes N-m} \otimes \mathbb{C}^{L+1}$ be small if and only if V outputs 1 with high probability on some proof $|\psi\rangle \in \mathcal{B}^{\otimes m}$. Suppose such a $|\psi\rangle$ exists. Then, for H_{in}, note that when one runs V on $|\psi\rangle$, the initial state should be $|\psi\rangle_p \otimes |0\rangle_a^{\otimes N-m}$, i.e. all ancilla qubits should be set to 0, with the purported proof in the proof register. But H_{in} enforces *precisely* this constraint for any $|\eta\rangle$. In particular, if the clock register of $|\eta\rangle$ is in state $|0\rangle\langle 0|_c$ and the ancilla register is *not* all zeroes, then we have $\langle\eta|H_{\text{in}}|\eta\rangle > 0$, i.e. $|\eta\rangle$ incurs an energy penalty. In other words, if $|\eta\rangle$ does not simulate the initial state of the verification procedure V, H_{in} penalizes $|\eta\rangle$. Next, for H_{out}, note that after running V on $|\psi\rangle$, we expect the first qubit in the proof register to be a 1 with high probability. Again, observe that H_{out} enforces exactly this constraint on $|\eta\rangle$ — if the clock register is in state $|L\rangle\langle L|_c$ and the first qubit reads 0, we again have $\langle\eta|H_{\text{out}}|\eta\rangle > 0$. Finally, H_{prop} follows the same idea by forcing $|\eta\rangle$ to encode in superposition a simulation of each step of the verification procedure V. It follows that the minimizing vector $|\eta\rangle$ is of the following form, often called a *history state*:

$$|\eta\rangle := \frac{1}{\sqrt{L+1}} \sum_{j=0}^{L} \left(V_j \dots V_1 |\psi\rangle_p \otimes |0\rangle_a^{\otimes N-m} \right) \otimes |j\rangle_c. \tag{1.87}$$

To recap, if there exists a $|\psi\rangle$ such that V accepts with high probability, then the history state $|\eta\rangle$ corresponds to a small eigenvalue of H. On the other hand, if no such $|\psi\rangle$ exists, either $|\eta\rangle$ will be of the form in Equation (1.87) (i.e. will faithfully simulate V), in which case we are hit with a large penalty by H_{out} since the answer qubit cannot be 1 with high probability, or $|\eta\rangle$ "cheats" by deviating from either the initial conditions or the intermediate steps of the protocol, in which case the terms H_{in} and H_{out} hit $|\eta\rangle$ with an energy penalty, respectively. Thus, the corresponding energy of $|\eta\rangle$ would be large. Of course, it remains to show that this intuition is indeed correct!

Before we begin, we first apply the following change of basis operator to H_{prop}, which greatly simplifies the analysis (intuition to follow):

$$W = \sum_{j=0}^{L} V_j \dots V_1 \otimes |j\rangle\langle j|_c. \tag{1.88}$$

Thus, instead of $|\eta\rangle$ and H, we consider $|\hat{\eta}\rangle := W|\eta\rangle$ and $\hat{H} := W^\dagger H W$. To see what $\hat{H}$ looks like, we analyze the action of W on each of H_{in}, H_{out}, and H_{prop} separately. Observe first that $\hat{H}_{\text{in}} := W^\dagger H_{\text{in}} W = H_{\text{in}}$, since at time 0, W implicitly applies the identity to the

proof and ancilla registers. Second, for H_{out}, we have

$$\hat{H}_{\text{out}} := W^\dagger H_{\text{out}} W = V^\dagger \left[(|0\rangle\langle 0| \otimes I_{\mathcal{B}^{\otimes m-1}})_p \otimes I_a \right] V \otimes |L\rangle\langle L|_c = (V^\dagger \otimes I_c) H_{\text{out}} (V \otimes I_c), \tag{1.89}$$

since at time L, W applies the entire circuit V. Finally, for H_{prop}, considering the effect of W on each component of H_j in Equation (1.86) separately and using simple algebra, one finds

$$W^\dagger H_j W = I_{p,a} \otimes \frac{1}{2}(|j-1\rangle\langle j-1| - |j-1\rangle\langle j| - |j\rangle\langle j-1| + |j\rangle\langle j|)_c = I_{p,a} \otimes \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}_c. \tag{1.90}$$

It follows that $\hat{H}_{\text{prop}} = \sum_j W^\dagger H_j W$ is tridiagonal and of the form

$$\hat{H}_{\text{prop}} = I_p \otimes I_a \otimes \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & \dots \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 & \dots \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & \dots \\ 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & \dots \\ 0 & 0 & 0 & -\frac{1}{2} & \ddots & \ddots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix} =: I_p \otimes I_a \otimes E_c, \tag{1.91}$$

where we have let E denote the tridiagonal matrix acting on the clock register for later reference. Intuitively, one can think of the change of basis W as "flushing out" the computation V, so that it is pushed to the very end to time step L (hence V only appears in H_{out}). This has the effect of simplifying $\hat{H}_{\text{prop}}$ to a nice tridiagonal form, since it no longer needs to keep track of the unitaries V_i.

Finally, observe that since W is unitary, $\hat{H}$ and H have precisely the same set of eigenvalues. We can thus work with $\hat{H}$ instead of H in our eigenvalue analysis. Hence, for the remainder of this section, by $|\eta\rangle$ we shall mean $|\hat{\eta}\rangle$, and by H, we mean $\hat{H}$. We now show that H has the correct spectral properties for both YES and NO instances of 5-LH.

YES case: H has a small eigenvalue

We have thus far set up a Hamiltonian $H \in \mathcal{H}(\mathcal{B}^{\otimes m} \otimes \mathcal{B}^{\otimes N-m} \otimes \mathbb{C}^{L+1})$ corresponding to the verification procedure $V \in \mathcal{U}(\mathcal{B}^{\otimes m} \otimes \mathcal{B}^{\otimes N-m})$. We now show that if there exists such a $|\psi\rangle \in \mathcal{B}^{\otimes m}$ which causes V to output 1 with high probability, then H must have a small eigenvalue.

Suppose there exists $|\psi\rangle$ such that a measurement of the first qubit of $V|\psi\rangle$ yields 1 with probability at least $1-\epsilon$. To demonstrate that H has a small eigenvalue, we explicitly construct a vector $|\eta\rangle \in \mathcal{B}^{\otimes m} \otimes \mathcal{B}^{\otimes N-m} \otimes \mathbb{C}^{L+1}$ such that $\langle\eta|H|\eta\rangle$ is small. Let

$$|\eta\rangle = |\psi\rangle_p \otimes |0\rangle_a^{\otimes N-m} \otimes |\gamma\rangle_c, \tag{1.92}$$

where

$$|\gamma\rangle := \frac{1}{\sqrt{L+1}} \sum_{j=0}^{L} |j\rangle. \tag{1.93}$$

We analyze $\langle\eta|H|\eta\rangle$ by considering H_{in}, H_{prop}, and H_{out} separately. First, observe that $\langle\eta|H_{\text{in}}|\eta\rangle = 0$, since the ancilla register of $|\eta\rangle$ is in the all zeroes state. For H_{prop}, we have that

$$\langle\eta|H_{\text{prop}}|\eta\rangle = \langle\eta|I_{p,a} \otimes E_c|\eta\rangle = \langle\gamma|E|\gamma\rangle = 0, \tag{1.94}$$

where in the last equality we have used the fact that the sum of each row and column of E is 0, implying $|\gamma\rangle$ is a 0-eigenvector of E. Note that we have not used the probability of V answering 1 yet — this now comes in handy for H_{out}, where

$$\begin{aligned} \langle\eta|H_{\text{out}}|\eta\rangle &= \langle\eta| \left(V^\dagger \left[(|0\rangle\langle 0| \otimes I_{\mathcal{B}^{\otimes m-1}})_p \otimes I_a \right] V \otimes |L\rangle\langle L|_c \right) |\eta\rangle \qquad (1.95) \\ &= \frac{1}{L+1} \left[\langle\psi|_p \otimes \langle 0|_a^{\otimes N-m} V^\dagger \right] \left[(|0\rangle\langle 0| \otimes I_{\mathcal{B}^{\otimes m-1}})_p \otimes I_a \right] \left[V|\psi\rangle_p \otimes |0\rangle_a^{\otimes N-m} \right]. \end{aligned}$$

Observe, however, that this expression corresponds to the probability that we begin with the proof $|\psi\rangle_p \otimes |0\rangle_a^{\otimes N-m}$, apply the verification V, and then measure the first qubit and obtain 0. By our assumption at the beginning of this section, this probability is at most ϵ. Hence,

$$\langle\eta|H_{\text{out}}|\eta\rangle \leq \frac{1}{L+1}\epsilon, \tag{1.96}$$

implying there must exist an eigenvalue for H of value at most $\epsilon/(L+1)$. Thus, if we have a YES-instance of our QMA problem P, then H has a small eigenvalue, as required.

NO case: H has no small eigenvalues

We now show that if there does not exist such a proof $|\psi\rangle$ which causes verification procedure V to output 1 with high probability, then H must have no small eigenvalues.

Suppose that for all proofs $|\psi\rangle$, V does not output 1 with probability more than ϵ. To lower bound the eigenvalues of H, we play a game of divide-and-conquer by letting

$H = A_1 + A_2$, where $A_1 := H_{\text{in}} + H_{\text{out}}$, and $A_2 := H_{\text{prop}}$, and analyzing the eigenvalues of A_1 and A_2 separately. The challenge arises in combining these separate eigenvalue estimates into eigenvalue estimates for H, since unfortunately, $[A_1, A_2] \neq 0$, implying that A_1 and A_2 do not diagonalize in a common basis. To surmount this obstacle, Kitaev uses the following approach [171]:

1. We first prove Kitaev's Geometric Lemma (Lemma 1.8), which takes as input operators B and C, as well as a set of parameters S dependent on B and C, and outputs a lower bound on the eigenvalues of $B + C$.

2. We compute the parameters S relevant to our specific operators A_1 and A_2, and plug them into Lemma 1.8 to show that $H = A_1 + A_2$ has no small eigenvalues.

We now state and prove Kitaev's Geometric Lemma.

Lemma 1.8 (Kitaev, Shen, Vyalyi [171], Geometric Lemma, Lemma 14.4)**.** *Let $A_1, A_2 \succeq 0$, such that the minimum* non-zero *eigenvalue of both operators is lower bounded by v. Assume that the null spaces $\mathcal{L}_1$ and $\mathcal{L}_2$ of A_1 and A_2, respectively, have trivial intersection, i.e. $\mathcal{L}_1 \cap \mathcal{L}_2 = \{\mathbf{0}\}$. Then*

$$A_1 + A_2 \succeq 2v \sin^2 \frac{\alpha(\mathcal{L}_1, \mathcal{L}_2)}{2} I \ , \tag{1.97}$$

where the angle *$\alpha(\mathcal{X}, \mathcal{Y})$ between $\mathcal{X}$ and $\mathcal{Y}$ is defined over unit vectors $|x\rangle$ and $|y\rangle$ as* $\cos\left[\angle(\mathcal{X}, \mathcal{Y})\right] := \max_{|x\rangle \in \mathcal{X}, |y\rangle \in \mathcal{Y}} |\langle x|y\rangle|$.

Note that if $\mathcal{X}$ and $\mathcal{Y}$ have non-trivial intersection, i.e. there exists $|x\rangle \neq \mathbf{0}$ such that $|x\rangle \in \mathcal{X}$ and $|x\rangle \in \mathcal{Y}$, then $\alpha(\mathcal{X}, \mathcal{Y})$ is trivially 0. Also, note that demanding $\mathcal{X} \cap \mathcal{Y} = \{\mathbf{0}\}$ is not equivalent to demanding $\mathcal{X}$ and $\mathcal{Y}$ be orthogonal — for example, the spaces $\mathcal{X} = \text{span}\,\{|0\rangle\}$ and $\mathcal{Y} = \text{span}\,\{|+\rangle\}$ contain elements which have non-zero overlap, but the sets have trivial intersection.

We now tackle step 1 of Kitaev's approach by proving Lemma 1.8.

Proof. By the definition of v, we have $A_1 \succeq v(I - \Pi_{\mathcal{L}_1})$ and $A_2 \succeq v(I - \Pi_{\mathcal{L}_2})$, where $\Pi_{\mathcal{X}}$ denotes the projector onto $\mathcal{X}$. Combining the latter two, it follows that it suffices to show $v(I - \Pi_{\mathcal{L}_1}) + v(I - \Pi_{\mathcal{L}_2}) \succeq 2v \sin^2(\alpha(\mathcal{L}_1, \mathcal{L}_2)/2)$. By rearranging terms and using the identity $\cos(2\theta) = 1 - 2\sin^2\theta$, this is the equivalent of showing

$$\Pi_{\mathcal{L}_1} + \Pi_{\mathcal{L}_2} \preceq \left[1 + \cos(\alpha(\mathcal{L}_1, \mathcal{L}_2))\right] I. \tag{1.98}$$

To upper bound the eigenvalues of $\Pi_{\mathcal{L}_1} + \Pi_{\mathcal{L}_2}$, suppose we have some eigenvector $|\zeta\rangle$ with corresponding eigenvalue $\lambda > 0$. Let $|x_1\rangle \in \mathcal{L}_1$ and $|x_2\rangle \in \mathcal{L}_2$ be unit vectors such that $\Pi_{\mathcal{L}_1}|\zeta\rangle = u_1|x_1\rangle$ and $\Pi_{\mathcal{L}_2}|\zeta\rangle = u_2|x_2\rangle$ for some real $u_1, u_2 > 0$. Then:

$$\lambda = \langle\zeta|(\Pi_{\mathcal{L}_1} + \Pi_{\mathcal{L}_2})|\zeta\rangle = u_1\langle\zeta|x_1\rangle + u_2\langle\zeta|x_2\rangle = u_1^2 + u_2^2. \tag{1.99}$$

Further, since $\lambda|\zeta\rangle = (\Pi_{\mathcal{L}_1} + \Pi_{\mathcal{L}_2})|\zeta\rangle = u_1|x_1\rangle + u_2|x_2\rangle$, we can also derive a non-equivalent expression for λ^2, i.e.

$$\lambda^2 = [\langle\zeta|\lambda][\lambda|\zeta\rangle] = (u_1\langle x_1| + u_2\langle x_2|)(u_1|x_1\rangle + u_2|x_2\rangle) = u_1^2 + u_2^2 + 2u_1u_2\,\mathrm{Re}\langle x_1|x_2\rangle, \tag{1.100}$$

where $\mathrm{Re}(x)$ denotes the real part of $x \in \mathbb{C}$. Combining Eqns. (1.99) and (1.100) by taking the following linear combination, we have:

$$\begin{aligned}
(1 + |\mathrm{Re}\langle x_1|x_2\rangle|)\lambda - \lambda^2 &= (1 + |\mathrm{Re}\langle x_1|x_2\rangle|)(u_1^2 + u_2^2) - (u_1^2 + u_2^2 + 2u_1u_2\,\mathrm{Re}\langle x_1|x_2\rangle) \\
&= u_1^2\,|\mathrm{Re}\langle x_1|x_2\rangle| + u_2^2\,|\mathrm{Re}\langle x_1|x_2\rangle| - 2u_1u_2\,\mathrm{Re}\langle x_1|x_2\rangle \qquad (1.101)\\
&= |\mathrm{Re}\langle x_1|x_2\rangle|\left(u_1^2 + u_2^2 \pm 2u_1u_2\right) \qquad (1.102)\\
&= |\mathrm{Re}\langle x_1|x_2\rangle|\,(u_1 \pm u_2)^2 \qquad (1.103)\\
&\geq 0. \qquad (1.104)
\end{aligned}$$

Moving λ^2 to the right side of the last inequality and dividing through by λ hence gives

$$\lambda \leq (1 + |\mathrm{Re}\langle x_1|x_2\rangle|) \leq 1 + \cos(\alpha(\mathcal{L}_1, \mathcal{L}_2)), \tag{1.105}$$

where the latter inequality follows straightforwardly from the definition of $\alpha(\mathcal{L}_1, \mathcal{L}_2)$. We thus have that all eigenvalues of $\Pi_{\mathcal{L}_1} + \Pi_{\mathcal{L}_2}$ are upper bounded by $1 + \cos(\alpha(\mathcal{L}_1, \mathcal{L}_2))$, which by Equation (1.98) implies the desired lower bound on $A_1 + A_2$. □

We now move to step 2 of Kitaev's approach, i.e. we now use Lemma 1.8 to lower bound the eigenvalues of $H = A_1 + A_2$. To do so, we must determine the values of parameters v and $\alpha(\mathcal{L}_1, \mathcal{L}_2)$ for A_1 and A_2 used in Lemma 1.8. Recall that Lemma 1.8 also requires $\mathcal{L}_1 \cap \mathcal{L}_2 = \{\mathbf{0}\}$ — we handle this constraint at the end of the section (at which point it will be obvious, given the analysis to come).

We start with v, which is the lower bound on the positive eigenvalues of both A_1 and A_2. Note that since $A_1 = H_{\text{in}} + H_{\text{out}}$ is simply a sum of commuting projectors, its eigenvalues must be non-negative integers. In particular, its smallest *positive* eigenvalue is at least 1. For A_2, since $A_2 = H_{\text{prop}} = I_{p,a} \otimes E_c$, its eigenvalues will be determined by those of E_c. The eigenvalues of the latter are [171] $\lambda_k = 1 - \cos[\pi k/(L+1)]$ for $0 \leq k \leq L$. This expression

is clearly minimized when $k = 1$ (note $k = 0$ would yield a zero eigenvalue), implying the smallest positive eigenvalue of A_2 is at least

$$1 - \cos(\pi/(L+1)) \geq c/L^2 \tag{1.106}$$

for some constant c. To see why this inequality holds, use the Taylor series expansion for $\cos x$ to show that whenever $x \leq 1$, one has

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots \leq 1 - \frac{x^2}{2!} + \frac{x^4}{4!} \leq 1 - \left(\frac{1}{2!} - \frac{1}{4!}\right) x^2 = 1 - cx^2. \tag{1.107}$$

Taking the minimum of our lower bounds for A_1 and A_2 thus yields that $v \in \Omega(1/L^2)$.

We next estimate the angle $\alpha(\mathcal{L}_1, \mathcal{L}_2)$ between the null spaces $\mathcal{L}_1$ and $\mathcal{L}_2$ of A_1 and A_2, respectively. This can be done by exploiting the structure of $\mathcal{L}_1$ and $\mathcal{L}_2$. In particular, we have that

$$\begin{aligned} \mathcal{L}_1 \quad = \quad & \left[(\mathcal{B}^{\otimes m})_p \otimes |0\rangle_a^{\otimes N-m} \otimes |0\rangle_c\right] \oplus \\ & \left[(\mathcal{B}^{\otimes N})_{p,a} \otimes \mathrm{span}(|1\rangle, \ldots |L-1\rangle)_c\right] \oplus \\ & \left[V^\dagger(|1\rangle \otimes \mathcal{B}^{\otimes N-1})_{p,a} \otimes |L\rangle_c\right], \end{aligned} \tag{1.108}$$

where each of the three terms in this expression follow directly from the definitions of H_{in} and H_{out} (e.g. any state with the clock register set to 0 and all zeroes in the ancilla is a 0-eigenvector of both H_{in} and H_{out}). Similarly, we have

$$\mathcal{L}_2 = (\mathcal{B}^{\otimes N})_{p,a} \otimes |\gamma\rangle_c, \tag{1.109}$$

which follows straightforwardly if we recall that $H_{\text{prop}} = I_{p,a} \otimes E_c$ and $E|\gamma\rangle = \mathbf{0}$, for $|\gamma\rangle$ defined in Equation (1.93).

To exploit this structure, instead of estimating $\alpha(\mathcal{L}_1, \mathcal{L}_2)$, we estimate $\cos^2 \alpha(\mathcal{L}_1, \mathcal{L}_2)$, which can be rewritten in the form (where the maximization is over unit vectors):

$$\cos^2 \alpha(\mathcal{L}_1, \mathcal{L}_2) = \max_{|x\rangle \in \mathcal{L}_1, |y\rangle \in \mathcal{L}_2} |\langle x|y\rangle|^2 = \max_{|x\rangle \in \mathcal{L}_1, |y\rangle \in \mathcal{L}_2} \langle y|x\rangle\langle x|y\rangle = \max_{|y\rangle \in \mathcal{L}_2} \langle y|\Pi_{\mathcal{L}_1}|y\rangle. \tag{1.110}$$

The last equality holds without loss of generality since the maximum for $\langle y|\Pi_{\mathcal{L}_1}|y\rangle$ is achieved by projecting onto a pure state $|x\rangle\langle x|$ for some $|x\rangle \in \mathcal{L}_1$. Let us upper bound the rightmost term in the equation above. Observe that by Equation (1.109), any $|y\rangle \in \mathcal{L}_2$ has the form $|y\rangle = |\zeta\rangle_{p,a} \otimes |\gamma\rangle_c$ for some $|\zeta\rangle \in \mathcal{B}^{\otimes m} \otimes \mathcal{B}^{\otimes N-m}$. Since by Equation (1.108), $\Pi_{\mathcal{L}_1}$ breaks down into a sum of three projections, we can bound $\langle y|\Pi_{\mathcal{L}_1}|y\rangle$ by determining the contribution of each projector separately when sandwiched by $|y\rangle$.

The contribution of the second projection is easiest to see — it is simply $(L-1)/(L+1)$, since every term in $|\gamma\rangle$ except $|0\rangle$ and $|L\rangle$ contribute $1/(L+1)$ to the sum.

As for the first and third projections, let $\mathcal{K}_1 = \mathcal{B}^{\otimes m} \otimes |0\rangle^{\otimes N-m}$ and $\mathcal{K}_2 = V^\dagger|1\rangle \otimes \mathcal{B}^{\otimes N-1}$. Then the contribution of the first and third projections is given by:

$$\langle y|(\Pi_{\mathcal{K}_1} \otimes |0\rangle\langle 0|_c + \Pi_{\mathcal{K}_2} \otimes |L\rangle\langle L|_c)|y\rangle = \frac{1}{L+1}\langle\zeta|(\Pi_{\mathcal{K}_1} + \Pi_{\mathcal{K}_2})|\zeta\rangle. \tag{1.111}$$

If we let $\varphi(\mathcal{K}_1, \mathcal{K}_2)$ denote the angle between $\mathcal{K}_1$ and $\mathcal{K}_2$, we can straightforwardly use Equation (1.98) to bound the quantity above by

$$\frac{1}{L+1}\langle\zeta|(\Pi_{\mathcal{K}_1} + \Pi_{\mathcal{K}_2})|\zeta\rangle \leq \frac{1}{L+1}\left(1 + \cos\varphi(\mathcal{K}_1, \mathcal{K}_2)\right). \tag{1.112}$$

Observe, however, that

$$\cos^2\varphi(\mathcal{K}_1, \mathcal{K}_2) = \max_{|k\rangle\in\mathcal{K}_1, |l\rangle\in\mathcal{K}_2} |\langle k|l\rangle|^2, \tag{1.113}$$

where $\mathcal{K}_1$ is just the set of initial states with all-zero ancilla for the verification procedure V, and $\mathcal{K}_2$ is the set of initial states for which applying V yields a 1 in the first qubit with certainty. Hence, the maximum overlap between vectors in $\mathcal{K}_1$ and $\mathcal{K}_2$ is directly tied to the maximum probability with which we can obtain outcome 1 with an initial state with all-zero ancilla. In particular, we have $\cos^2\varphi(\mathcal{K}_1, \mathcal{K}_2)$ equals the maximum probability of outputting 1. Since in this section we are dealing with the NO case, however, meaning no proof can cause an output of 1 with probability greater than ϵ, we have $\cos^2\varphi(\mathcal{K}_1, \mathcal{K}_2) \leq \epsilon$, implying:

$$\frac{1}{L+1}\left(1 + \cos\varphi(\mathcal{K}_1, \mathcal{K}_2)\right) \leq \frac{1}{L+1}\left(1 + \sqrt{\epsilon}\right). \tag{1.114}$$

Adding the contributions of all three projections thus yields:

$$\cos^2\alpha(\mathcal{L}_1, \mathcal{L}_2) = \max_{|y\rangle\in\mathcal{L}_2}\langle y|\Pi_{\mathcal{L}_1}|y\rangle \leq \left(\frac{L-1}{L+1}\right) + \left(\frac{1+\sqrt{\epsilon}}{L+1}\right) = 1 - \frac{1-\sqrt{\epsilon}}{L+1}. \tag{1.115}$$

Using the identity $\sin^2 x + \cos^2 x = 1$, this implies $\sin^2\alpha(\mathcal{L}_1, \mathcal{L}_2) \geq (1-\sqrt{\epsilon})/(L+1)$. Then, since $\sin^2\frac{x}{2} \geq \frac{1}{4}\sin^2 x$ (shown using the identity $\sin(2x) = 2\sin x\cos x$), we have

$$\sin^2\frac{\alpha(\mathcal{L}_1, \mathcal{L}_2)}{2} \geq \frac{1}{4}\sin^2\alpha(\mathcal{L}_1, \mathcal{L}_2) \geq \frac{1-\sqrt{\epsilon}}{4(L+1)}. \tag{1.116}$$

Finally, we have all estimates required to use Lemma 1.8: $v = \Delta/L^2$ for some constant Δ and $\sin^2[\alpha(\mathcal{L}_1, \mathcal{L}_2)/2] \geq (1-\sqrt{\epsilon})/[4(L+1)]$. In addition, given Equations (1.108)

and (1.109), it is now easy to see that $\mathcal{L}_1 \cap \mathcal{L}_2 = \{\mathbf{0}\}$ (as required by Lemma 1.8), since any state of the tensor product form $|\psi\rangle_{p,a} \otimes |\gamma\rangle_c$ cannot live in $\mathcal{L}_1$. Plugging everything into Lemma 1.8, we conclude that in the NO case, the minimum eigenvalue of H is of the order $\Omega((1-\sqrt{\epsilon})/L^3)$ (i.e. H has no "small" eigenvalues). As required by Definition 1.7, note that this lower bound is inverse polynomially separated from the upper bound on the smallest eigenvalue of H from the YES case if we first apply error reduction to V to bring ϵ inverse polynomially close to 0.

Is the Hamiltonian H 5-local?

We have so far set up a Hamiltonian H whose eigenvalues are small or large, depending on whether we have a YES or NO instance of our QMA problem P, respectively. We now ask: Is H 5-local?

The answer is *almost*. Recall that $H \in \mathcal{H}(\mathcal{B}^{\otimes m} \otimes \mathcal{B}^{\otimes N-m} \otimes \mathbb{C}^{L+1})$, where the counter register is $\mathbb{C}^{L+1}$. If we implement the counter straightforwardly using $O(\log L)$ qubits, the resulting operations on it, such as incrementing the counter, could require updating all $O(\log L)$ qubits, making H $(\log L)$-local at best. In order to circumvent this, Kitaev [171] uses a different representation for the counter for which any operation requires acting on at most 3 qubits of the counter. Specifically, we let H act on $\mathcal{B}^{\otimes m} \otimes \mathcal{B}^{\otimes N-m} \otimes \mathcal{B}^{L}$, where the counter register is now given in *unary*, i.e. $|j\rangle \in \mathbb{C}^{L+1}$ is represented as

$$|\underbrace{1,\ldots,1}_{j},0,\ldots,0\rangle. \tag{1.117}$$

The operator basis $|i\rangle\langle j|$ for $\mathcal{L}(\mathbb{C}^{L+1})$ translates to this new representation as follows. Operator $|j\rangle\langle j| \in \mathcal{L}(\mathbb{C}^{L+1})$ is mapped to $|1\rangle\langle 1|_j \otimes |0\rangle\langle 0|_{j+1}$ in the new space, i.e. being in state $|j\rangle$ in the old encoding is equivalent to having the jth qubit set to 1 and the $(j+1)$-th qubit set to 0 in the new encoding. Similarly, operator $|j-1\rangle\langle j|$ is mapped to $|1\rangle\langle 1|_{j-1} \otimes |0\rangle\langle 1|_j \otimes |0\rangle\langle 0|_{j+1}$, i.e. if we think of $|j-1\rangle\langle j|$ as moving us from state $|j\rangle$ to $|j-1\rangle$, this is equivalent in the new encoding to flipping the jth bit to 0, followed by a safety check that qubits $j-1$ and $j+1$ are 1 and 0, respectively. The remaining basis elements are defined analogously. These operations are at most 3-local. Combined with the fact that H is based on the verification circuit V, which itself is composed of 2-qubit unitaries V_i, we have that H is 5-local Hamiltonian, as desired.

With H being 5-local, there is one final issue to be addressed — since the counter is now represented using a larger space, one must deal with the possibility of *invalid* settings

to the counter register. To discourage such behavior, a fourth penalty term is added to H acting only on the counter space, namely

$$H_{\text{stab}} := I_{p,a} \otimes \sum_{j=1}^{L-1} |0\rangle\langle 0|_j \otimes |1\rangle\langle 1|_{j+1}. \tag{1.118}$$

Hence, the new H is given by $H = H_{\text{in}} + H_{\text{prop}} + H_{\text{out}} + H_{\text{stab}}$. Note that H_{stab} discourages counter states which are not of the form in Equation (1.117), i.e. states containing the subsequence 01 are given an energy penalty.

Does the previous analysis of the smallest eigenvalue of H still hold when H_{stab} is added to the picture? The answer is *yes*. The YES case is easy to see, since all valid counter states are in the null space of H_{stab}. Thus, an honest proof receives no energy penalty from H_{stab}, as desired.

For the NO case, let $\mathcal{S} = \mathcal{B}^{\otimes m} \otimes \mathcal{B}^{\otimes N-m} \otimes \mathbb{C}^{L+1}$ (the original space we had defined H as acting on). Observe that $H_{\text{in}} + H_{\text{prop}} + H_{\text{out}}$ and H_{stab} both act invariantly on $\mathcal{S}$, meaning they map operators in $\mathcal{S}$ to operators in $\mathcal{S}$. Thus, we can split our analysis into two independent cases: when H acts on $\mathcal{S}$, and when H acts on the orthogonal complement of $\mathcal{S}$, denoted $\mathcal{S}^{\perp}$. In the former case, H_{stab} is just the zero operator with respect to $\mathcal{S}$; thus, the previous eigenvalue analysis goes through unscathed, yielding an eigenvalue lower bound on H of $\Omega((1-\sqrt{\epsilon})/L^3)$. As for the second case when H is restricted to $\mathcal{S}^{\perp}$, observe that H_{stab} always administers an energy penalty, since $\mathcal{S}^{\perp}$ contains only invalid counter states. Since H_{stab} is a sum of commuting projectors, its eigenvalues will be non-negative integers — in particular, its smallest non-zero eigenvalue is at least 1. Since $H_{\text{in}} + H_{\text{prop}} + H_{\text{out}} \succeq 0$, it follows that when restricted to $\mathcal{S}^{\perp}$, we have $H \succeq 1$. Taking the minimum of the estimates for the two cases of $\mathcal{S}$ and $\mathcal{S}^{\perp}$ yields the desired bound that the smallest eigenvalue of H is still in $\Omega((1-\sqrt{\epsilon})/L^3)$, despite the new representation for the counter. This concludes Kitaev's proof that 5-local Hamiltonian is complete for QMA.

1.6 Quantum correlations

As mentioned earlier, the growing field of quantum computation and information has positively impacted both computer science and physics. The next area this thesis studies has in particular benefited greatly from this cross-fertilization, and is the study of *quantum correlations*. Here, we are interested in understanding correlations between individual quantum subsystems of a larger composite system. Specifically, we shall introduce and discuss two notions of quantum correlations: *quantum entanglement* and *non-classical correlations*.

Motivation. We mention two reasons why the study of quantum correlations is important. The first is that the existence of certain correlations predicted by quantum theory, specifically quantum entanglement, has long troubled physicists. In a letter to Max Born in 1947, for example, Einstein dubs entanglement as "spukhafte Fernwirkung", or "spooky action at a distance" [50]. This mentality was moreover the basis for the rejection of quantum mechanics as a complete physical theory a decade earlier by the famous Einstein, Podolsky, and Rosen (EPR) paper of 1935 [89]. Thus, a better understanding of quantum correlations appears to be key to understanding both the nature of our world around us, as well as our theories describing this world. The second reason is that quantum correlations are generally believed to be required for quantum computers to outperform their classical counterparts. It has been rigorously shown, for example, that in the pure-state setting, the amount of entanglement present in a quantum system must grow with the problem size if a quantum computation is to achieve an exponential speedup over classical computers [160]. Thus, a better understanding of quantum correlations may prove advantageous for designing quantum algorithms, as well as for uncovering the boundary between classical and quantum computing.

1.6.1 Quantum entanglement

The canonical notion of quantum correlations between quantum systems dates back to the EPR paper of 1935 [89], and is called *quantum entanglement*. The name "entanglement" was coined by physicist Erwin Schrödinger, who used the term "Vershränkung" in 1935 [221], which in colloquial "non-physicist" German means "folding of the arms" [61]. Much has been discovered in the field of *entanglement theory* over the last two decades, from its quantification and characterization, to its manipulation and use for quantum computational and information theoretic tasks. In particular, what was once considered "spooky action at a distance" is now regarded as a valuable resource in quantum information (see, e.g. [151]). In this thesis, entanglement is not a primary focus, but rather has important connections to non-classical correlations in the results of Chapters 7 and 8. We give a brief introduction to entanglement here; the reader is referred to the surveys of Bruß [61] and Horodecki$^{\otimes 4}$ [151] for further details.

To begin, the canonical example of an entangled state is the two-qubit EPR pair,

$$\left|\phi^+\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \tag{1.119}$$

By observing that $\mathrm{Tr}_1(|\phi^+\rangle\langle\phi^+|) = \mathrm{Tr}_2(|\phi^+\rangle\langle\phi^+|) = I/2$, we have one of the characteristic traits of quantum mechanics — that for quantum systems, knowledge of the whole quantum

system does not imply knowledge of its parts. Since entangled (pure) states, such as the EPR pair, cannot be written as a *product state* $|\psi_1\rangle \otimes |\psi_2\rangle$ of single qubit states $|\psi_1\rangle, |\psi_2\rangle$, a primary area of study in quantum information has been the quantification of "how far" an entangled state is from product form. (Note that all classical states, by which we mean bit strings, are of product form.)

The answer to this question varies greatly depending on context. For bipartite *pure* states $|\psi_{AB}\rangle \in \mathcal{D}(\mathbb{C}^m \otimes \mathbb{C}^n)$, the canonical measure of entanglement is given by the *entropy of entanglement* [151],

$$E(|\psi_{AB}\rangle) = S(\mathrm{Tr}_B |\psi_{AB}\rangle\langle\psi_{AB}|) = S(\mathrm{Tr}_A |\psi_{AB}\rangle\langle\psi_{AB}|), \tag{1.120}$$

where $S(\rho) := -\mathrm{Tr}(\rho \log(\rho))$ is the von Neumann entropy of ρ. It holds that $0 \leq E(|\psi_{AB}\rangle) \leq \log(\min(m, n))$, where the lower bound is achieved if and only if a state is of product form, and the upper bound is achieved if and only if a state is *maximally entangled*, such as the EPR pair.

The definition of $E(|\psi_{AB}\rangle)$ is perhaps better motivated by the fact that any bipartite $|\psi_{AB}\rangle \in \mathbb{C}^m \otimes \mathbb{C}^n$ can be written in terms of the *Schmidt decomposition*, such that

$$|\psi_{AB}\rangle = \sum_{i=1}^{\min(m,n)} \alpha_i |\psi_i\rangle \otimes |\phi_i\rangle. \tag{1.121}$$

Here, the real $\alpha_i \geq 0$ are called *Schmidt coefficients*, and the sets $\{|\psi_i\rangle\}$ and $\{|\phi_i\rangle\}$ are orthonormal bases for $\mathbb{C}^m$ and $\mathbb{C}^n$, respectively, known as the *Schmidt bases*. The Schmidt decomposition is extremely useful in quantum information; some of our results in Chapter 2, for example, depend heavily on it. A proof of existence for the Schmidt decomposition is straightforward, and makes use of the vec mapping (defined in the proof of Corollary 7.8 here) and singular value decomposition for operators; we refer the reader to [246] for details. To now see the connection between E and the Schmidt decomposition, let $\mathbf{p} \in \mathbb{R}^{\min(m,n)}$ with $\mathbf{p}(i) = \alpha_i^2$. Then, $E(|\psi\rangle) = H(\mathbf{p})$, where $H(\mathbf{p}) := -\sum_i p(i) \log p(i)$ is the Shannon entropy of probability distribution $\mathbf{p}$. In the other words, the more "tightly concentrated" the Schmidt coefficients of $|\psi_{AB}\rangle$ are, the less entangled $|\psi_{AB}\rangle$ is. Note that a state is product if and only if it has a Schmidt coefficient $\alpha_i = 1$, and a state is maximally entangled if and only if all its Schmidt coefficients are $1/\sqrt{d}$ for $d = \min(m, n)$.

Moving to the mixed state case, the quantification of entanglement becomes much more complex. Most generally, we say operator $\rho \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ is *separable* (i.e. unentangled) if and only if it can be written [250]

$$\rho = \sum_i A_i \otimes B_i \tag{1.122}$$

for $A_i \in \mathrm{Pos}(\mathcal{X})$ and $B_i \in \mathrm{Pos}(\mathcal{Y})$. This definition of separability (with the added trace one constraint) was first given by Werner [254]. We denote the set of separable operators acting on $\mathcal{X} \otimes \mathcal{Y}$ as $\mathrm{Sep}(\mathcal{X}, \mathcal{Y})$. Note that $\mathrm{Sep}(\mathcal{X}, \mathcal{Y})$ is a convex cone; this property is vital to the results of Section 4.5. Here, a *cone* is a set $S \subseteq \mathcal{X}$ such that $\lambda x \in S$ for all $x \in S$ and all $\lambda \geq 0$. If additionally $u + v \in S$ for $u, v \in S$, then S is called a *convex cone*. When we restrict ourselves to the set of separable *density* operators in $\mathrm{Sep}(\mathcal{X}, \mathcal{Y})$ (i.e. we impose the trace one constraint), we obtain a *convex set*. (A set $S \subset \mathcal{X}$ is called *convex* if $px + (1-p)y \in S$ for all $x, y \in S$ and $0 \leq p \leq 1$.) The set of separable density operators has the following properties, which prove useful in Section 4.5: It is compact and contains a ball around the maximally mixed state (which is, of course, separable) [124, 125, 126].

The problem of determining whether a given density operator $\rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{Y})$ is in $\mathrm{Sep}(\mathcal{X}, \mathcal{Y})$ (where one is allowed to work in time polynomial in the dimension), known as the *Quantum Separability Problem*, was shown NP-hard to solve within inverse exponential precision by Gurvits [123] (see also the work of Ioannou [154]). This was later extended to inverse polynomial precision by the present author [105], and shortly thereafter independently by Beigi [39]. Recently, a breakthrough result of Christandl, Brandão, and Yard [53] has shown that the problem is *quasi-polynomial*-time solvable for the case of *constant* precision; the result goes via a powerful new de Finetti-type theorem for the Frobenius (and LOCC, where LOCC stands for *local operations and classical correlations*) norms.

Thus, as suggested by the NP-hardness of Quantum Separability Problem, in the mixed-state case there is no known efficient test for separability, unlike the pure-state case. To this end, there have been many mixed state entanglement measures proposed to date; the reader is referred to the survey of Horodecki$^{\otimes 4}$ [151] for an in-depth look.

Here, we mention two entanglement detection schemes used in this thesis. The first is the popular approach proposed by Peres [206] known as the *positive partial transpose* (PPT) test, which plays a role in Chapters 7 and 8. Specifically, consider the super-operator $I \otimes T$ acting on space $\mathcal{L}(\mathcal{X} \otimes \mathcal{Y})$, where T denotes the transpose map. Then, given any $\rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{Y})$, if $(I \otimes T)(\rho) \not\succeq 0$, then ρ is not separable. This follows since for any separable operator $\sum_i A_i \otimes B_i$,

$$(I \otimes T)\left(\sum_i A_i \otimes B_i\right) = \sum_i A_i \otimes T(B_i) \succeq 0. \tag{1.123}$$

Above, we have used the fact that the transpose map does not change the spectrum of an operator. The PPT test is known to be necessary and sufficient for pure states of all dimensions, and for mixed states of (2×2) and (2×3)-dimensional systems [206, 146]. In higher dimensions, however, we remark that there exist mixed entangled states which nevertheless

have a positive partial transpose; such states are called *bound entangled* [147, 144]. Bound entangled states have the property that they cannot be *distilled*, meaning roughly that in the asymptotic limit, given many copies of a bound entangled state ρ, there does not exist an LOCC (local operations and classical communication) protocol which can extract the entanglement present in the copies of ρ into pure EPR pairs. The quantification of just how much entanglement can be distilled in this sense is given by another entanglement measure, the *distillable entanglement* [210]; this makes a brief appearance in Chapter 7.

Finally, there is an easy way to compute the partial transpose given a matrix representation of state $\rho \in \mathcal{D}(\mathbb{C}^m \otimes \mathbb{C}^n)$: Namely, partition the matrix into $(m \times n)$-dimensional blocks, and take the tranpose of each block individually. For example, for the EPR pair $|\phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, we have

$$(I \otimes T)(\left|\phi^+\right\rangle\!\left\langle\phi^+\right|) = (I \otimes T) \begin{pmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{pmatrix} \not\geq 0. \tag{1.124}$$

The second entanglement detection scheme we define here is the *relative entropy of entanglement* [238, 145]. Specifically, define for $\rho, \sigma \in \mathcal{D}(\mathcal{X})$ the *relative entropy* as

$$S(\rho||\sigma) := -\mathrm{Tr}(\rho \log \sigma) - S(\rho). \tag{1.125}$$

Then, for $\rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{Y})$, the relative entropy of entanglement is defined as

$$E_R(\rho) = \min_{\sigma \in \mathrm{Sep}(\mathcal{X},\mathcal{Y})} S(\rho||\sigma). \tag{1.126}$$

The following properties regarding E_R hold [238]: It takes value 0 if and only if $\rho \in \mathrm{Sep}(\mathcal{X}, \mathcal{Y})$, is invariant under local unitary operations, is convex, reduces to the entropy of entanglement for pure states, and is an upper bound on the distillable entanglement (see also [212]). It is further non-increasing under LOCC, which follows since $S(\rho||\sigma) \geq S(\Phi(\rho)||\Phi(\sigma))$ for any TPCP map Φ [237]. In fact, a stronger and physically more relevant statement holds — that even if we allow *post-selection* after performing an LOCC measurement, the value of E_R does not increase on average [238]. In other words, let $\{K_i\}$ be a complete set of Kraus operators for a TPCP map, i.e. $\sum_i K_i^\dagger K_i = I$. Then, letting $\rho_i := K_i \rho K_i^\dagger$, it holds that

$$E_R(\rho) \geq \sum_i \mathrm{Tr}(\rho_i) E_R\left(\frac{\rho_i}{\mathrm{Tr}(\rho_i)}\right). \tag{1.127}$$

We close this section by noting that the definition of separability presented here extends straightforwardly to the multipartite setting. The structure of multipartite entanglement, however, is markedly more daunting than in the bipartite case.

1.6.2 Non-classical correlations

Having discussed quantum entanglement, we now turn our attention to another form of quantum correlations, called simply *non-classical correlations*. Such correlations have attracted much attention in the last decade or so, both in terms of their characterization and quantification, as well as with respect to their use as a resource in quantum information. We begin by motivating the study of non-classical correlations, and follow with definitions. We then discuss the role of such correlations in quantum information processing tasks, and close by surveying a number of known non-classicality measures. The reader is referred to Modi *et al.* [195] for a more comprehensive survey of the topic.

Motivation. As mentioned earlier, it is known that in the case of pure-state quantum computation, entanglement is a necessary resource for exponential speedup over classical computers [160]. What happens, however, if we instead consider *mixed*-state quantum computing? This is a particularly relevant question, as typically one deals with mixed states in a laboratory setting due to noise from the environment. In 1998, Knill and Laflamme [174] proposed a model of computing known as *Deterministic Quantum Computing with one clean qubit (DQC1)* (see Chapter 5), wherein all but one qubit of the computation are initialized to the maximally mixed state — in other words, the quantum computation acts on a highly mixed state. (Note that this model is motivated experimentally by nuclear-magnetic resonance (NMR) information processing, in which states are highly mixed.) Yet, this model can perform the task of (normalized) trace estimation of a given unitary exponentially faster than the best known classical algorithm. This raises the question: Is entanglement also the root of the believed speedup in DQC1? (This is a natural question since "very highly mixed" states are separable due to a ball around the maximally mixed state in the set of separable quantum states [124, 125, 126].) Or are there other correlations possibly at play? Recent work has suggested that although Erwin Schrödinger once wrote that entanglement is *"not just one of many traits, but the characteristic trait of quantum physics"* [221] (as quoted in [195]), that between purely classical correlations and entanglement, there lies another form of quantum correlations whose nature is only now beginning to be understood. Such correlations are known simply as *non-classical correlations*.

Defining non-classical correlations. We now define what we mean by non-classical correlations. To begin, we say that a quantum state $\rho_{AB} \in \mathcal{D}(\mathcal{X} \otimes \mathcal{Y})$, henceforth denoted as ρ to avoid clutter, is *strictly classically correlated* or *classical* if it can be diagonalized in a local product basis. In other words, ρ is classical if there exist local orthonormal bases $\{|\psi_i\rangle\}, \{|\phi_i\rangle\}$ for $\mathcal{X}$ and $\mathcal{Y}$, respectively, such that

$$\rho = \sum_{ij} \lambda_i |\psi_i\rangle\langle\psi_i| \otimes |\phi_j\rangle\langle\phi_j|, \tag{1.128}$$

for $\{\lambda_i\}$ the eigenvalues of ρ. Note that such a state is simply an embedding of a classical bipartite distribution into the quantum formalism. Any state not satisfying this definition is called *non-classical*. We remark that this definition of classicality extends straightforwardly to the multipartite setting.

Continuing in the bipartite setting, a particularly interesting class of states which subsume the classical states are the so-called *classical-quantum (CQ)* states, which are only classical in system A. Specifically, a state $\rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{Y})$ is CQ if there exists a local orthonormal basis $\{|\psi_i\rangle\}$ for $\mathcal{X}$ such that

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \otimes \rho_i, \tag{1.129}$$

for $\{p_i\}$ a probability distribution and for arbitrary $\rho_i \in \mathcal{D}(\mathcal{Y})$. Note that system A in ρ simply plays the role of a classical label: Upon measuring it in basis $\{|\psi_i\rangle\}$ and obtaining outcome i, we know the induced state ρ_i in B. Also, observe that CQ states are separable. An analogous definition straightforwardly yields the similar class of quantum-classical (QC) states. As an aside, note that neither classical nor CQ states form a convex set, unlike the set of separable quantum states.

Non-classical correlations and quantum information processing. A number of connections are known between non-classical correlations and quantum information processing tasks, involving for example local broadcasting [209, 188], extended state merging [62], the locking of classical correlations [87, 78, 259, 49] (see Chapter 5), assisted optimal state discrimination [213, 181], remote state preparation [74], entanglement distribution [230, 68], and activation of non-classical correlations into entanglement [208] (see Chapter 7, and also related work by Streltsov, Kampermann and Bruß [231]). We now discuss two of these tasks: local broadcasting and entanglement distribution.

We begin with the task of *local broadcasting*. Specifically, generalizing the no-cloning theorem of Section 1.4.5 is the following statement. Given a state $\rho \in \mathcal{D}(\mathcal{X})$, we say

$\rho' \in \mathcal{D}(\mathcal{X} \otimes \mathcal{X})$ is a *broadcast state* for ρ if

$$\mathrm{Tr}_1(\rho') = \mathrm{Tr}_2(\rho') = \rho, \tag{1.130}$$

where the $1:2$ split is across the two copies of $\mathcal{X}$. Now, suppose we are given a set of density operators $\{\rho_i\} \subseteq \mathcal{D}(\mathcal{X})$, and some arbitrary starting state σ. Then, the statement we are interested in is that there exists a TPCP map $\Lambda \in T(\mathcal{X} \otimes \mathcal{X})$ which, for all i, achieves the mapping $\rho_i \otimes \sigma \mapsto \rho_i' \in \mathcal{D}(\mathcal{X} \otimes \mathcal{X})$ for ρ_i' a broadcast state for ρ_i if and only if the ρ_i pairwise commute. This is called the *no-broadcasting theorem* [37, 36]. With respect to non-classical correlations, a variant of this theorem is the *no-local-broadcasting* theorem of Piani *et al.* [209, 188], which states that for any bipartite state $\rho_{AB} \in \mathcal{D}(\mathcal{X} \otimes \mathcal{Y})$, there exist local TPCP maps $\Theta_A \in T(\mathcal{X}, \mathcal{X} \otimes \mathcal{X})$ and $\Theta_B \in T(\mathcal{Y}, \mathcal{Y} \otimes \mathcal{Y})$ such that $\Theta_A \otimes \Theta_B(\rho_{AB})$ is a broadcast state if and only if ρ_{AB} is strictly classical. Thus, the classicality of correlations in ρ is strongly tied to how well one can carry out the information theoretic task of local broadcasting.

We next discuss the task of entanglement distribution. Consider a tripartite system ABC consisting of Alice, Bob, and a carrier system C. Roughly, the goal of entanglement distribution is for Alice and Bob to increase the entanglement between their systems A and B by having Alice send Bob the carrier system C. More specifically, we imagine Alice holds systems A and C to start, and Bob holds system B. Alice applies some encoding operation jointly to A and C. She then sends C to Bob. Bob finally applies some decoding operation to B and C. We now ask: Is the entanglement in the $AC : B$ cut before the protocol was run strictly smaller than the entanglement in the $A : BC$ cut after Bob receives the carrier C? What is perhaps most surprising about this task is that the answer to this question can be yes even if the carrier C is *not* entangled with A and B throughout the protocol [73]! Motivated by the question of whether non-classical correlations could be the resource behind this phenomenon, Streltsov *et al.* [230] and Chuan *et al.* [68] (both works appeared concurrently and independently) showed that (definitions to follow)

$$\left| E_R^{AC|B}(\rho_{ABC}) - E_R^{A|BC}(\sigma_{ABC}) \right| \leq \delta_R^{AB|C}(\sigma_{ABC}), \tag{1.131}$$

where ρ_{ABC} is the state before the protocol is run, σ_{ABC} is the state once Bob receives C from Alice, and where we measure non-classicality by the relative entropy of discord (RED) $\delta_R^{AB|C}$ of Equation (1.140) (to be defined shortly) across the $AB : C$ cut, and we measure entanglement by the relative entropy of entanglement $E_R^{AC|B}$ ($E_R^{A|BC}$) across the $AC : B$ ($A : BC$) cut . In other words, the amount of entanglement which can be transferred from Alice to Bob is bounded by the amount of non-classical correlations between the carrier C and AB (after Alice has applied her encoding operation). Note thus that this upper bound

can be non-zero even if C is unentangled with A and B throughout the protocol (and in fact must be non-zero for the example of Cubitt *et al.* [73] mentioned above).

Quantifying non-classical correlations. Finally, we close this section by discussing a number of known non-classicality measures.

The formal notion of CQ states first arose with the works of Ollivier and Zurek [203] and Henderson and Vedral [138], where a measure of quantum correlations dubbed the *quantum discord* was proposed. The aim of this measure is to quantify purely *quantum* correlations in a bipartite state ρ. To define the discord, recall first that the (classical) *mutual information* is a measure of correlation between (classical) random variables A and B, i.e.

$$\mathcal{I}(A:B) = H(A) + H(B) - H(A,B), \tag{1.132}$$

where H is the Shannon entropy defined in Section 1.6.1 and $H(A,B) = -\sum_{a,b} \Pr(A = a \cap B = b) \log \Pr(A = a \cap B = b)$. Using the fact that $\Pr(B|A) = \Pr(A \cap B)/\Pr(A)$, one can straightforwardly also express the mutual information as

$$\mathcal{J}(A:B) = H(B) - H(B|A), \tag{1.133}$$

where $H(B|A) := \sum_a \Pr(A = a) H(B|A = a)$. Although $\mathcal{I}$ and $\mathcal{J}$ are equivalent in the classical setting, their quantum counterparts no longer share the same relationship. Specifically, the quantum mutual information can be defined as

$$\mathcal{I}(\rho_{AB}) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}), \tag{1.134}$$

where recall $\rho_A = \mathrm{Tr}_B(\rho_{AB})$. However, a quantum variant of $\mathcal{J}$ is non-trivial to define, since it requires specifying a value for B for the conditional entropy $H(A|B)$ — in particular, unlike the classical setting, quantumly the choice of measurement basis is non-trivial. To this end, for rank-one projective measurement $\left\{\Pi_j^A\right\}$, one defines [203] a quantum conditional entropy

$$S\left(\rho_{B|\left\{\Pi_j^A\right\}}\right) := \sum_j p_j S\left((\Pi_j^A \otimes I^B)\rho(\Pi_j^A \otimes I^B)\Big/ p_j\right), \tag{1.135}$$

where $p_j = \mathrm{Tr}(\Pi_j^A \otimes I^B \rho)$. Then, a quantum version of $\mathcal{J}$ for given measurement basis $\left\{\Pi_j^A\right\}$ can be defined as

$$\mathcal{J}_{\left\{\Pi_j^A\right\}}(\rho) = S(\rho_B) - S\left(\rho_{B|\left\{\Pi_j^A\right\}}\right). \tag{1.136}$$

Note that $\mathcal{J}_{\{\Pi_j^A\}}(\rho)$ quantifies the amount of classical correlations which can be extracted from ρ_{AB} via a projective measurement on one party; since we are in the end interested in purely quantum correlations, intuitively one would thus choose the *optimum* measurement $\{\Pi_j^A\}$ here so as to extract all purely classical correlations, leaving only quantum correlations behind. With this in mind, the quantum discord is now defined as

$$\delta(\rho) := \mathcal{I}(\rho) - \max_{\{\Pi_j^A\}} \mathcal{J}_{\{\Pi_j^A\}}(\rho) = S(\rho_A) - S(\rho_{AB}) + \min_{\{\Pi_j^A\}} S\left(\rho_{B|\{\Pi_j^A\}}\right). \tag{1.137}$$

The discord is [195] non-negative, non-symmetric with respect to exchange of systems A and B, invariant under local unitaries, and most importantly for our discussion here, takes value zero if and only if ρ is CQ [203, 79]. Moreover, there exist *separable* states, such as the two-qubit state

$$\frac{1}{2}|0\rangle\langle 0| \otimes |0\rangle\langle 0| + \frac{1}{2}|+\rangle\langle +| \otimes |1\rangle\langle 1|, \tag{1.138}$$

which have non-zero discord, thus showing that discord quantifies correlations beyond entanglement. (Aside: The state above is studied further in Chapters 6, 7, and 8.)

The next measure of non-classical correlations we discuss is the *geometric quantum discord* [75]. Let $\mathcal{CQ} \subseteq \mathcal{D}(\mathcal{X} \otimes \mathcal{Y})$ denote the set of classical-quantum states. Then for $\rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{Y})$ the geometric discord is defined as

$$\delta_G(\rho) := \min_{\sigma \in \mathcal{CQ}} \| \rho - \sigma \|_{\mathrm{F}}^2 = \min_{\{\Pi_j^A\}} \left\| \rho - \sum_j \Pi_j^A \rho \Pi_j^A \right\|_{\mathrm{F}}^2, \tag{1.139}$$

where $\| \cdot \|_{\mathrm{F}}$ is the Frobenius norm and the second equality was shown by Luo and Fu [186]. The name *geometric* derives from the fact that the measure attempts to quantify distance from $\mathcal{CQ}$ via a metric. We have included the right-most expression in Equation (1.139) as it offers another intuitive interpretation of non-classical correlations involving *disturbance under measurement*. Namely, recall that in the classical world, there always exists a choice of measurement basis $\{\Pi_j^A\}$ (the computational basis) leaving the target state undisturbed. In the quantum setting, however, this is in general not the case. For example, this is an intuitive reason why CQ states are considered classical in A; there exists a measurement basis acting invariantly on A. The second expression for $\delta_G(\rho)$ in Equation (1.139) thus attempts to understand how much ρ must be disturbed in a (rank one projective) measurement, regardless of the choice of local measurement basis for A.

The next non-classicality measure we discuss is similar to the geometric discord, but replaces the Frobenius norm with the *relative entropy*. We thus arrive at the *relative*

entropy of discord (RED) [196],

$$\delta_R(\rho) = \min_{\sigma \in \mathcal{CQ}} S(\rho||\sigma). \tag{1.140}$$

An analogous definition for the case of general strictly classically correlated states goes under the name of the *relative entropy of quantumness (REQ)* [54, 187, 118, 217, 196]; this is studied further in Chapters 7 and 8.

Interestingly, the RED turns out to be equal to (a variant of) another measure of non-classical correlations we discuss next, the *quantum deficit* [145]. The latter's definition is motivated by work extraction from quantum systems coupled to a heat bath. Roughly, the idea here is that a state is strictly classically correlated if and only if the same amount of work can be drawn from the global state versus from the local subsystems after allowing a suitably restricted subset of local operations and classical communication (LOCC) known as *closed LOCC*. The variant of the deficit which is equal [145] to the RED is the *one-way deficit* $\Delta^{\rightarrow}$, given by (simplified from the original definition):

$$\Delta^{\rightarrow} := \min_{\{\Pi_j^A\}} S\left(\sum_j \Pi_j^A \rho \Pi_j^A\right) - S(\rho_{AB}). \tag{1.141}$$

Here, $\{\Pi_j^A\}$ again denotes a rank-one projective measurement. The correspondence between RED and the deficit does not stop here, however; the two-sided analogue of the RED, the REQ, is equal [145] to the so-called *zero-way deficit* $\Delta^{\emptyset}$:

$$\Delta^{\emptyset} := \min_{\{\Pi_i^A\},\{\Pi_j^B\}} S\left(\sum_{ij} \Pi_i^A \otimes \Pi_j^B \rho \Pi_i^A \otimes \Pi_j^B\right) - S(\rho_{AB}). \tag{1.142}$$

We have discussed a number of non-classicality measures here. Later in Chapter 6, we introduce a novel measure of non-classical correlations based on local unitary operations, which for $(2 \times N)$-dimensional quantum states turns out to coincide with the geometric discord. Chapters 7 and 8 then introduce and study a protocol for "activating" non-classical correlations into entanglement, while also providing an operational interpretation for the REQ.

Chapter 2

Approximation algorithms for QMA-complete problems

This chapter is based on [108]:

S. Gharibian and J. Kempe. Approximation algorithms for QMA-complete problems. In *Proceedings of 26th IEEE Conference on Computational Complexity*, pages 178-188, 2011, DOI: 10.1109/CCC.2011.15, © 2011 IEEE, ieeexplore.ieee.org.

Approximation algorithms for classical constraint satisfaction problems are one of the main research areas in theoretical computer science. In this chapter, we define a natural approximation version of the QMA-complete local Hamiltonian problem and initiate its study. We present two main results. The first shows that a non-trivial approximation ratio can be obtained in the class NP using product states. The second result (which builds on the first one), gives a polynomial time (classical) algorithm providing a similar approximation ratio for dense instances of the problem. The latter result is based on an adaptation of the "exhaustive sampling method" by Arora *et al.* [28] to the quantum setting, and might be of independent interest.

2.1 Introduction and results

In the last few years, the quantum analog of the class NP, the class QMA [171], has been extensively studied, and several QMA-complete problems have been found [182, 55,

184, 40, 215, 158, 223, 253]. Arguably the most important (and historically first) QMA-complete problem is the k-local Hamiltonian problem [171, 164, 202, 163, 20]. Recall from Section 1.5.4 that here, the input is a set of Hamiltonians (Hermitian matrices), each acting on at most k-qubits each. The task is to determine the largest eigenvalue of the sum of these Hamiltonians. This problem generalizes the central NP-hard problem MAX-k-CSP, where we are given a set of Boolean constraints on k variables each, with the goal to satisfy as many constraints as possible. The local Hamiltonian problem is of significant interest to complexity theorists and to physicists studying properties of physical systems alike (e.g. [60, 15, 58, 17, 69, 176, 222]).

Moving to the classical scenario, the theory of NP-completeness is one of the great success stories of classical computational complexity [27]. It was soon realized that many natural optimization problems are NP-hard, and are hence unlikely to have polynomial time algorithms. A natural question (both in theory and in practice) is to look for polynomial time algorithms that produce solutions that are close to optimum. More precisely, one says that an algorithm achieves an *approximation ratio* of $c \in [0, 1]$ for a certain maximization problem if on all inputs, the value of the algorithm's output is at least c times that of the optimum solution (the output value should also be at most the optimal solution). The closer c is to 1, the better the approximation. The investigation of approximation algorithms is, after decades of heavy research, still a very active area (e.g., [141, 236]). For many central NP-hard problems, tight polynomial time approximation algorithms are known.

In the context of QMA-complete problems, it is thus natural to search for approximation algorithms for these problems, and in particular for the local Hamiltonian problem. The question we address here is: *How well can one efficiently approximate the k-local Hamiltonian problem?*

It should be noted that a large host of heuristics has been developed in the physics community to approximate properties of local Hamiltonian systems (see, e.g., [69] for a survey) and this area is extremely important in the study of physical systems. However, the systematic complexity theoretic study of approximation algorithms for QMA-complete problems is still very much in its infancy, and our work is one of the first steps in this research direction. We note that there has been a lot of interest in recent years [17, 6] in establishing a so-called quantum PCP theorem [30, 29], which amounts to showing that for some constant $c < 1$ close enough to 1, approximating the k-local Hamiltonian (or related problems) to within c is QMA-hard. Our results can also be seen as a natural continuation of that investigation.

Our results: Let us start by precisely defining the optimization version of the local Hamiltonian problem, which is parameterized by two integers k and d, which we always think of as constants. Note that the definition below differs slightly from that given in Section 1.5.4, Definition 1.7; we discuss the differences after stating the definition.

Definition 2.1 (MAX-k-local Hamiltonian problem on d-level systems (qudits))**.** *An instance of the problem consists of a collection of* $\binom{n}{k}$ *Hermitian matrices, one for each subset of* k *qudits. The matrix* $H_{i_1,\ldots,i_k}$ *corresponding to some* $1 \leq i_1 \leq \cdots \leq i_k \leq n$ *is assumed to act on those qudits (terms acting on less than* k *qudits can be incorporated by tensoring them with the identity), to be positive semidefinite, and to have operator norm at most* 1*. We call any pure or mixed state* ρ *on* n *qudits an* assignment *and define its* value *to be* $\mathrm{Tr}(H\rho)$ *where* $H = \sum_{i_1,\ldots,i_k} H_{i_1,\ldots,i_k}$*. The goal is to find the largest eigenvalue of* H *(denoted* OPT*), or equivalently, the maximum value obtained by an assignment. We say that an algorithm provides an* approximation ratio *of* $c \in [0,1]$ *if for all instances, it outputs a value that is between* $c \cdot \mathrm{OPT}$ *and* OPT*.*

This definition, we believe, is the natural quantum analog of the MAX-k-CSP problem. We note that it differs slightly from the usual definition of the k-local Hamiltonian problem. Namely, we consider maximization (as opposed to minimization), and also restrict the terms of H to be positive semidefinite, and have norm at most 1 (the latter two contraints are also common to Definition 1.7; more generally, the local terms of H can be arbitrary Hermitian operators). As long as one considers the *exact* problem, these assumptions are without loss of generality, and do not affect the definition, as seen by simply scaling the Hamiltonians and adding multiples of identity as necessary. However, when dealing with the *approximation* version, these assumptions are important for the problem to make sense; for instance, one cannot meaningfully talk about approximation ratios if the optimum can take both negative and positive values. That is why we require the terms to be positive semidefinite. The requirement that the terms have operator norm at most 1 does not affect the problem and later allows us to conveniently define dense instances. Finally, changing the maximization to a minimization would lead to an entirely different approximation problem: the quantum analogue of MIN-CSP (e.g. [167]). Minimization problems are, generally speaking, harder than maximization problems, and we leave this research direction for future work.

Before stating our results, we state a trivial way to get a d^{-k}-approximation for MAX-k-local Hamiltonian. Observe that the maximally mixed state has at least d^{-k} overlap with the reduced density matrix of the optimal assignment on any k particles. A similar property holds classically, where a random assignment gives (in expectation) a d^{-k} approximation of MAX-k-CSP. We now describe our two main results.

Approximation by product states. One inherently quantum property of the local Hamiltonian problem is the fact that the optimal state might in general be highly entangled (and hence not efficiently describable in polynomial time or space). This is why we do not require outputting the assignment itself in the above definition. If, however, the optimal assignment (or some other good assignment) was guaranteed to be a *product state*, then we could describe it efficiently. The following theorem shows just that.

Theorem 2.2. *For an instance of MAX-k-local Hamiltonian with optimal value* OPT*, there is a (pure) product state assignment that has value at least* OPT/d^{k-1}*.*

This result is *tight* for product states in the case of 2-local Hamiltonians (we remark that 2-local Hamiltonians are often the most relevant case from a physics perspective). For example, consider the Hamiltonian on 2-qubits that projects onto the EPR state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. It is easy to see that no product state achieves value more than $1/2$. For general d and k, we can only show that product states cannot achieve an approximation ratio greater than $1/d^{\lfloor k/2 \rfloor}$ (see Section 2.2, where better bounds in more specific cases are also discussed).

If we could efficiently find the best product state assignment, we would obtain an algorithm achieving a non-trivial d^{-k+1} approximation ratio. Unfortunately, this problem is NP-complete, since it would allow one to solve (e.g.) the special case of MAX-k-SAT (as discussed in Section 1.5.4, for each clause C acting on variables $\{i_1, \ldots, i_k\}$ in an instance of MAX-k-SAT, define the corresponding Hamiltonian term $H_{i_1,\ldots,i_k}$ diagonal in the computational basis and projecting onto the satisfying assignments for C. Then, without loss of generality, the optimal product state assignment can be taken to be a computational basis state), implying such an algorithm cannot exist unless P = NP. Still, the theorem has the following interesting implication: It shows that unless NP = QMA, approximating the local Hamiltonian problem to within a factor less than d^{-k+1} is not QMA-hard. This follows simply because product states have polynomial size classical descriptions. (More accurately, since one uses a polynomial number of classical bits to approximately specify a product state in NP, the ratio in the implication above is $d^{-k+1} - f(m)$ for some function f which scales inverse exponentially in the input size m.)

A polynomial time approximation algorithm for dense instances. Our second result gives a classical polynomial time approximation algorithm for *dense* instances of the local Hamiltonian problem. This result is perhaps our technically most challenging one, and we hope the techniques we develop might turn out useful elsewhere.

Dense instances of classical constraint satisfaction problems have been studied in depth [81, 101, 115, 28, 82, 23, 38, 83]. Our result is inspired by work of Arora *et al.* [28] who provide a polynomial time approximation scheme, or PTAS (i.e., an efficient $1-\varepsilon$ approximation algorithm for any fixed $\varepsilon > 0$), for several types of dense constraint satisfaction problems. In the classical case, dense (for 2-local constraints) simply means that the average degree in the constraint graph is $\Omega(n)$, or equivalently, that the optimum is $\Omega(n^2)$. In analogy, we define an instance of MAX-k-local Hamiltonian to be *dense* if $\mathrm{OPT} = \Omega(n^k)$, or equivalently, if $\mathrm{Tr}(H\frac{I}{d^n}) = \Omega(n^k)$ (the equivalence follows from the fact that the mixed state assignment I/d^n has value between OPT and OPT/d^k).

It is not hard to see that the (exact) dense local Hamiltonian problem remains QMA-hard (see Section 2.3.3). We hope the dense case might be of practical interest to physicists who study systems of particles by incorporating all possible interactions between them. Our second main result is the following:

Theorem 2.3. *For all $\varepsilon > 0$ there is a polynomial time $(1/d^{k-1} - \varepsilon)$-approximation algorithm for the dense MAX-k-local Hamiltonian problem over qudits.*

Theorem 2.3 follows immediately by combining Theorem 2.2 with the following theorem, which gives an approximation scheme for the problem of optimizing over the set of product states.

Theorem 2.4. *Let OPT_P denote the value of the optimal product state assignment for an instance of MAX-k-local Hamiltonian H. Then, for all $\varepsilon > 0$, there is a polynomial time algorithm which outputs a product state assignment attaining value at least $\mathrm{OPT}_P - \epsilon n^k$. For all $\varepsilon > 0$, this yields an efficient $(1-\epsilon)$-approximation algorithm for computing OPT_P for dense MAX-k-local Hamiltonian.*

We remark that the algorithm of Theorem 2.4 also applies in the *minimization* setting, in which one is interested in computing the *smallest* eigenvalue of k-local Hamiltonian H. Here, our algorithm outputs a value at most $\mathrm{OPT}_P + \epsilon n^k$.

Proof ideas and new tools: The proofs of Theorem 2.2 and Theorem 2.4 are independent and employ different techniques. To show the product state approximation guarantee, we show a slightly stronger statement: For *any* assignment $|\Psi\rangle$, there is a way to construct a product assignment of at least d^{-k+1} its value. The proof is constructive (given $|\Psi\rangle$): we use a type of recursive Schmidt decomposition of $|\Psi\rangle$ to obtain a mixture of product states whose value is off by at most the desired approximation factor (see Section 2.2).

Our second result is technically more challenging and introduces a few new ideas to this problem, inspired by work of Arora *et al.* [28] in the classical setting. We illustrate the main ideas for MAX-2-local Hamiltonian on n qubits. Recall that our goal is to find a PTAS for the local Hamiltonian problem *over product states*. The value of the optimal product state assignment, OPT_P, can be written

$$\mathrm{OPT}_P \quad = \quad \max \quad \sum_{i=1}^{n} \sum_{j \in N(i)} \mathrm{Tr}(H_{i,j}(\rho_i \otimes \rho_j)) \quad \text{s.t.} \quad \rho_i \succeq 0 \text{ and } \mathrm{Tr}(\rho_i) = 1 \quad \text{for } 1 \leq i \leq n, \tag{2.1}$$

where $N(i)$ is the set of indices j for which a local Hamiltonian term $H_{i,j}$ is present. We might call this a *quadratic* semidefinite program, as the maximization is quadratic in the ρ_i (and as such not efficiently solvable in general). Note, however, that if the terms in the maximization were *linear*, then we would obtain a semidefinite program (SDP), which is efficiently solvable [121]. To "linearize" our optimization, we use the "exhaustive sampling method" developed by Arora *et al.* [28] (a method which was later key in many developments in property testing, e.g. [115]). We write each Hamiltonian term in a basis that separates its two qubits, for instance the Pauli basis $\{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}$, $H_{i,j} = \sum_{k,l=0}^{3} \alpha_{kl}^{ij} \sigma_k \otimes \sigma_l$. For $i = 1, \ldots, n$ and $k = 0, 1, 2, 3$, define

$$c_k^i := \sum_{j \in N(i)} \sum_{l=0}^{3} \alpha_{kl}^{ij} \mathrm{Tr}(\sigma_l \rho_j). \tag{2.2}$$

If we knew the values of c_k^i for the optimal ρ_i, then solving the SDP below would yield the optimal ρ_i:

$$\begin{aligned} \max \quad \sum_{i=1}^{n} \sum_{k=0}^{3} c_k^i \mathrm{Tr}(\sigma_k \rho_i) \quad \text{s.t.} \quad \rho_i \succeq 0 \text{ and } \mathrm{Tr}(\rho_i) = 1 \qquad & \text{for } 1 \leq i \leq n, \\ \sum_{j \in N(i)} \sum_{l=0}^{3} \alpha_{kl}^{ij} \mathrm{Tr}(\sigma_l \rho_j) = c_k^i \quad & \text{for } 1 \leq i \leq n \text{ and } 0 \leq k \leq 3. \end{aligned} \tag{2.3}$$

Of course, this reasoning is circular, as in order to obtain the c_k^i we need the optimal ρ_i. The crucial idea is now to use *sampling* to *estimate* the c_k^i. More precisely, assume for a second that we could sample $O(\log n)$ of the ρ_i randomly from the optimal assignment. Then, by standard sampling bounds, with high probability over the choice of the sampled qubits we can estimate all the c_k^i to within an additive error $\pm \varepsilon n$ for some ε. If we had these estimates a_k^i for the c_k^i, we could solve the SDP above with the slight modification

that the last constraint should be $a_k^i - \varepsilon n \leq \sum_{j \in N(i)} \sum_l \alpha_{kl}^{ij} \mathrm{Tr}(\sigma_l \rho_j) \leq a_k^i + \varepsilon n$. With high probability over the sampled qubits, this SDP will give a solution that is within an additive εn^2 of the optimal one (more subtle technicalities and all calculations can be found in Section 2.3). Moreover, it is possible to derandomize the sampling procedure to obtain a deterministic algorithm (Section 2.3.3).

Of course, we are still in the realm of wishful thinking, because in order to sample from the optimal solution, we would need to know it, which is precisely what we set out to do. However, the number of qubits we wish to sample is only *logarithmic* in the input size. Thus, to simulate the sampling procedure, we can pick a random subset of $O(\log n)$ qubits, and simply *iterate* through all possible assignments on them (with an appropriate δ-net over the density matrices, which incurs a small additional error) in polynomial time! Our algorithm then runs the SDP for each iteration, and we are guaranteed that at least one iteration will return a solution within εn^2 of the optimal one. Because the denseness assumption guarantees that OPT_P is $\Omega(n^2)$, our additive approximation turns into a factor $(1-\varepsilon)$-approximation, as desired. All details, the runtime of the algorithm and error bounds for the general k-local case on qudits are given in Section 2.3. We remark that the approach above works analogously in the setting where the objective function involves minimization instead of maximization.

Previous and related work: We note that many heuristics have been developed in the physics community to approximate properties of local Hamiltonian systems and this area is extremely important in the study of physical systems (e.g. [255, 256, 205, 214, 220, 207, 69, 204]). Our focus here is, however, on *rigorous* bounds (unlike a heuristic) on the approximation guarantee of algorithms for the *general* problem (we allow interactions of arbitrary types occurring on arbitrary graphs, in contrast to the more common approach of studying specific local Hamiltonian models with certain classes of allowed interactions). In this area, to our knowledge, few results are known. In the setting of *relative*-error approximation, as studied here, the first and only previous result we are aware of is that of Bansal, Bravyi and Terhal [35], who give a PTAS for a special case of the local Hamiltonian problem, so called quantum Ising spin glasses, for the case where the instance is on a planar graph and of bounded degree. Roughly, this PTAS is obtained by dividing the graph into constant size chunks, which can be solved directly, and ignoring the constraints between chunks (this incurs an error proportional to the number of such constraints, which is small because the graph is planar). In the setting of *absolute*-error approximation, in 1D models, rigorous results such as Hasting's 1D area law are known for gapped systems [133] (where it is also shown that the ground state is well-approximated by a Matrix Product State [240]), and rigorous approximation methods are known for 1D [16, 222] and for 2-local

Hamiltonians on qubits where the two-qubit interaction strengths are weak [57]. Finally, we remark that the use of a product state ansatz is closely related to the mean-field approximation or Hartree-Fock method in physics (see, e.g. [90]).

Discussion and open questions: Our two results give approximations to the local Hamiltonian problem. Although at first glance, our approximation ratio of $1/d^{k-1}$ may appear an incremental improvement over the trivial random assignment strategy, there are three important notes that should be kept in mind: The first is that many classical NP-hard problems, such as MAX-3-SAT (a special case of MAX-k-CSP where each constraint is the disjunction ("OR") of k variables or their negation), are *approximation resistant* (e.g. [132, 32]), meaning that unless P=NP, there do not even exist non-trivial approximation ratios beyond the random assignment strategy. For example, for MAX-3-SAT it is NP-hard to do better than the approximation ratio of 7/8 achieved by random assignment [131]. Thus, showing the existence of a non-trivial approximation ratio is typically a big step in the classical setting. Moreover, it could have been conceivable that for MAX-k-local Hamiltonian, analogously to MAX-3-SAT, outperforming the random assignment strategy would have been *QMA-hard*. Yet our results show that unless NP=QMA, this is not the case. The second important note that should be kept in mind is that our work considers the local Hamiltonian problem in its full generality by allowing arbitrary constraints on an arbitrary interaction graph. It could be (and is the case, for example, in [35]) that for more restricted classes of local Hamiltonian models, better approximation ratios are achievable. Third, the currently *best* approximation algorithm for MAX-k-CSP gives an approximation ratio of only about $0.44k/2^k$ for $k > 2$ [63] (for $k = 2$, one can achieve 0.874 [180]. See also the work of Raghavendra [211]) and this is, moreover, essentially the best possible under a plausible complexity theoretic conjecture (namely, the Unique Games Conjecture [168]) [234, 130, 218, 32]. This is to be contrasted with our $2/2^k$-approximation ratio for the case of $d = 2$ (i.e. qubit systems), which we show can be achieved by product state assignments for *arbitrary* (i.e. even non-dense) MAX-k-local Hamiltonian instances (in the non-dense case, however, we do not show how to *efficiently* find a product state achieving this ratio). This raises the important open question: Is our approximation ratio tight?

Our product state approximation shows that approximating the local Hamiltonian problem to within d^{-k+1} is in NP. It would be interesting to know if this approximation ratio could also be achieved in polynomial time. If not, it might lead to an intriguing state of affairs where for low approximation ratios the problem is efficiently solvable, for medium ratios it is in NP but not efficiently solvable, and for high ratios it is QMA-hard (assuming a quantum PCP theorem exists). Further, as mentioned earlier, our work can be viewed as

negative progress towards a quantum PCP theorem in that, by Theorem 2.2, a quantum PCP theorem with hardness ratio $c \leq d^{-k+1}$ cannot exist unless NP=QMA.

To obtain our results for the case of dense local Hamiltonians, we have introduced the exhaustive sampling technique of Arora *et al.* [28] to the setting of low-degree semidefinite programs. We linearize such programs using exhaustive sampling in combination with a careful analysis of the error coming from working with δ-nets on density matrices. We remark that it seems we cannot simply apply the results of [28] for *smooth Polynomial Integer Programs* as a black-box to our setting. This is due to our aforementioned need for a δ-net, as well as the requirement that our assignment be a positive semidefinite operator. We address the latter issue by extending the techniques of [28] to the realm of positive semidefinite programs by introducing the notion of "degree-k inner products" over Hermitian operators to generalize the concept of degree-k polynomials over real numbers, and performing the more complex analysis that ensues. We hope that this technique will be of much wider applicability, particularly considering the growing use of semidefinite programs in numerous areas of quantum computing and information (e.g. [88, 155, 178]).

Another open question is whether similar ideas can be used to approximate other QMA-complete problems, such as the Consistency problem [182]. Moreover, can we obtain polynomial time algorithms without the denseness assumption? And are there special cases of the local Hamiltonian problem for which there is a PTAS (other than for planar Ising spin glasses [35])? Of course, we do not expect a PTAS for all instances of the local Hamiltonian problem, as this would contradict known hardness results for special classical cases of the problem. However, perhaps there exist other classes of physically relevant instances of the problem for which a PTAS does exist. Finally, can our scheme be extended to work with more general classes of quantum assignments than product states, such as Matrix Product States [240]?

Organization of this chapter: In Section 2.2, we prove our result on product state approximations (Theorem 2.9 and the ensuing proof of Theorem 2.2), show its tightness in the 2-local case and provide the upper bound of $d^{-\lfloor k/2 \rfloor}$ for the best possible approximation by product states. Section 2.3 gives our polynomial time approximation algorithm and develops the general sampling and SDP-based technique we use. It also shows that the dense local Hamiltonian problem remains QMA-complete. As some of the proofs and notation of Section 2.3 are rather technical, we have deferred the full proofs of this section to Section 2.4 in order to facilitate reading.

2.2 Product states yield a $1/d^{k-1}$-approximation for qudits

We now show that product state assignments achieve a non-trivial approximation ratio for MAX-k-local Hamiltonian, i.e. Theorem 2.2. To do so, we first define the *recursive Schmidt decomposition* (RSD, Definition 2.5) of a state $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$, and for ease of exposition, the corresponding notion of a *Schmidt cut* (Definition 2.6). We then state and prove the key to our approach, the *Mixing Lemma* (Lemma 2.7), which shows how to use the RSD to eliminate the entanglement across a particular Schmidt cut of $|\psi\rangle$ while maintaining the desired approximation ratio. Lemma 2.8 and Theorem 2.9 then expand on this by showing how to apply the Mixing Lemma to multiple Schmidt cuts. From Theorem 2.9, a proof of Theorem 2.2 easily follows. We close with a discussion of the tightness of the approximation ratio given by Theorem 2.2.

We first define the terms Recursive Schmidt Decomposition and Schmidt cut.

Definition 2.5 (Recursive Schmidt Decomposition (RSD))**.** Given a state $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$, we define its *recursive Schmidt decomposition* as the expression obtained by recursively applying the Schmidt decomposition on each qudit from 1 to $n-1$ inclusive. More formally, we define the RSD of $|\psi\rangle$ as follows:

- (Base case) If $n = 1$, then $\mathrm{RSD}(|\psi\rangle) = |\psi\rangle$.
- (Recursive case) If $n > 1$, then $\mathrm{RSD}(|\psi\rangle) = \sum_{i=1}^{d} \alpha_i |\psi_i\rangle \otimes \mathrm{RSD}(|\phi_i\rangle)$, where $|\psi_i\rangle \in \mathbb{C}^d$, $|\phi_i\rangle \in (\mathbb{C}^d)^{\otimes n-1}$, $\sum_{i=1}^{d} \alpha_i^2 = 1$, $\{|\psi_i\rangle\}$ is an orthonormal basis for the first qudit of $|\psi\rangle$, and $\{|\phi_i\rangle\}$ is a set of orthonormal vectors for the remaining $n-1$ qudits of $|\psi\rangle$.

(This definition is relative to some fixed ordering of the qudits. The specific choice of ordering is unimportant in our scenario, as any decomposition output by such a process suffices to prove Theorem 2.2.) For example, the RSD for 3-qubit $|\psi\rangle$ is

$$|\psi\rangle = \alpha_1 |a_1\rangle \otimes (\beta_1 |b_1\rangle |c_1\rangle + \beta_2 |b_2\rangle |c_2\rangle) + \alpha_2 |a_2\rangle \otimes (\beta'_1 |b'_1\rangle |c'_1\rangle + \beta'_2 |b'_2\rangle |c'_2\rangle), \tag{2.4}$$

for $\alpha_1^2 + \alpha_2^2 = \beta_1^2 + \beta_2^2 = \beta_1'^2 + \beta_2'^2 = 1$, $\{|a_i\rangle\}_i$ an orthonormal basis for qubit 1, $\{|b_i\rangle\}_i$ and $\{|b'_i\rangle\}_i$ orthonormal bases for qubit 2, and $\{|c_i\rangle\}_i$ and $\{|c'_i\rangle\}_i$ orthonormal bases for qubit 3.

Definition 2.6 (Schmidt cut)**.** For any $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$ with Schmidt decomposition $|\psi\rangle = \sum_{i=1}^{d} \alpha_i |w_i\rangle |v_i\rangle$, where $\alpha_i \in \mathbb{R}$ with $\sum_i \alpha_i^2 = 1$, $|w_i\rangle \in \mathbb{C}^d$ and $|v_i\rangle \in (\mathbb{C}^d)^{\otimes n-1}$, and for

any $|\phi\rangle \in (\mathbb{C}^d)^{\otimes m}$, we refer to the expansion $|\phi\rangle \otimes \left(\sum_{i=1}^d \alpha_i |w_i\rangle|v_i\rangle\right)$ as the *Schmidt cut* at qudit $m+1$. We say that a projector Π *crosses* this Schmidt cut if Π acts on qudit $m+1$ and at least one qudit $i \in \{m+2, \ldots, m+n\}$.

The heart of our approach is the following Mixing Lemma, which provides, for *any* assignment $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$, an explicit construction through which the entanglement across the first Schmidt cut of $|\psi\rangle$ can be eliminated, while maintaining at least a $(1/d)$-approximation ratio relative to the value $|\psi\rangle$ achieves against any local Hamiltonian $H \in \mathcal{H}((\mathbb{C}^d)^{\otimes n})$.

Lemma 2.7 (Mixing Lemma)**.** *Given state $|\psi\rangle$ on n qudits with Schmidt cut on qudit 1 given by $|\psi\rangle = \sum_{i=1}^d \alpha_i |w_i\rangle|v_i\rangle$, where $\alpha_i \in \mathbb{R}$ with $\sum_i \alpha_i^2 = 1$, $|w_i\rangle \in \mathbb{C}^d$ and $|v_i\rangle \in (\mathbb{C}^d)^{\otimes n-1}$, define $\rho := \sum_{i=1}^d \alpha_i^2 |w_i\rangle\langle w_i| \otimes |v_i\rangle\langle v_i|$. Then, given projector Π acting on some subset $\mathcal{S}$ of the qudits, if Π crosses the Schmidt cut, then $\mathrm{Tr}(\Pi\rho) \geq \frac{1}{d}\mathrm{Tr}(\Pi|\psi\rangle\langle\psi|)$. Otherwise, $\mathrm{Tr}(\Pi\rho) = \mathrm{Tr}(\Pi|\psi\rangle\langle\psi|)$.*

Proof. Case 2 follows easily by noting that the given Schmidt decomposition of $|\psi\rangle$ implies $\mathrm{Tr}_1(\rho) = \mathrm{Tr}_1(|\psi\rangle\langle\psi|)$ and $\mathrm{Tr}_{2,\ldots,n}(\rho) = \mathrm{Tr}_{2,\ldots,n}(|\psi\rangle\langle\psi|)$. To prove case 1, we observe by straightforward expansion that

$$\mathrm{Tr}(\Pi|\psi\rangle\langle\psi|) = \mathrm{Tr}(\Pi\rho) + \sum_{i<j} \alpha_i\alpha_j\langle w_i|\langle v_i|\Pi|w_j\rangle|v_j\rangle + \alpha_i\alpha_j\langle w_j|\langle v_j|\Pi|w_i\rangle|v_i\rangle. \tag{2.5}$$

Then, by defining for each i vector $|a_i\rangle := \alpha_i\Pi|w_i\rangle|v_i\rangle$, we have

$$\sum_{i<j} \alpha_i\alpha_j\langle w_i|\langle v_i|\Pi|w_j\rangle|v_j\rangle + \alpha_i\alpha_j\langle w_j|\langle v_j|\Pi|w_i\rangle|v_i\rangle = \sum_{i<j}\langle a_i|a_j\rangle + \langle a_j|a_i\rangle, \tag{2.6}$$

since $\Pi^2 = \Pi$. Applying the fact that $\langle a|b\rangle + \langle b|a\rangle \leq \| |a\rangle \|_2^2 + \| |b\rangle \|_2^2$ for $|a\rangle, |b\rangle \in (\mathbb{C}^d)^{\otimes n}$ thus implies

$$\sum_{i<j}\langle a_i|a_j\rangle+\langle a_j|a_i\rangle \leq \sum_{i<j} \| |a_i\rangle \|_2^2+\| |a_j\rangle \|_2^2 = (d-1)\sum_i \alpha_i^2\langle w_i|\langle v_i|\Pi|w_i\rangle|v_i\rangle = (d-1)\mathrm{Tr}(\Pi\rho), \tag{2.7}$$

from which the claim follows. □

The following simple extension of Lemma 2.7 simplifies our proof of Theorem 2.9.

Corollary 2.8. *Define $|\psi'\rangle := |\phi\rangle \otimes |\psi\rangle$, where $|\phi\rangle \in (\mathbb{C}^d)^{\otimes m}$ for $m > 0$ and $|\psi\rangle$ is defined as in Lemma 2.7, and let $\rho \in \mathcal{D}(\mathbb{C}^d)^{\otimes n}$ be obtained from $|\psi\rangle$ as in Lemma 2.7. Then, for any projector Π acting on a subset $\mathcal{S}$ of the qudits, if Π crosses the Schmidt cut of $|\psi'\rangle$ at qudit $m+1$, we have $\mathrm{Tr}(\Pi|\phi\rangle\langle\phi| \otimes \rho) \geq \frac{1}{d}\mathrm{Tr}(\Pi|\psi'\rangle\langle\psi'|)$. Otherwise, $\mathrm{Tr}(\Pi|\phi\rangle\langle\phi| \otimes \rho) = \mathrm{Tr}(\Pi|\psi'\rangle\langle\psi'|)$.*

Proof. Immediate by applying the proof of Lemma 2.7 with the following modifications: (1) Define $|a_i\rangle := \alpha_i \Pi |\phi\rangle |w_i\rangle |v_i\rangle$, and (2) if $\mathcal{S} \subseteq \{1, \ldots, m\} \cup \{m+2, \ldots, m+n\}$ (i.e. this is one of two ways for Π not to cross the cut — the other way is for $\mathcal{S} \subseteq \{1, \ldots, m+1\}$), observe that by the same arguments as in Lemma 2.7 for case 2 and the product structure between $|\phi\rangle$ and $|\psi\rangle$ in $|\psi'\rangle$ that $\mathrm{Tr}_{m+1}(|\phi\rangle\langle\phi| \otimes \rho) = \mathrm{Tr}_{m+1}(|\psi'\rangle\langle\psi'|)$. □

Lemma 2.7 shows that the state ρ obtained by *mixing* the d Schmidt vectors of $|\psi\rangle$, as opposed to taking their *superposition*, suffices to achieve a $(1/d)$-approximation across the first Schmidt cut. By iterating this argument over *all* $n-1$ Schmidt cuts, we now prove that a mixture of all (product) states appearing in the RSD of $|\psi\rangle$ achieves an approximation ratio of $1/d^{k-1}$.

Theorem 2.9. *For any n-qudit assignment $|\psi\rangle$ with RSD $|\psi\rangle = \sum_{i=1}^{d^{n-1}} \sqrt{p_i} |\phi_i\rangle$, where $\sum_i p_i = 1$ and $\{|\phi_i\rangle\}_{i=1}^{d^{n-1}}$ is a set of orthonormal product vectors in $(\mathbb{C}^d)^{\otimes n}$, define $\rho := \sum_{i=1}^{d^{n-1}} p_i |\phi_i\rangle\langle\phi_i|$. Then, for any projector Π acting on some subset $\mathcal{S} \subseteq \{1, \ldots, n\}$ of qudits with $|\mathcal{S}| = k$, we have $\mathrm{Tr}(\Pi\rho) \geq \frac{1}{d^{k-1}} \mathrm{Tr}(\Pi |\psi\rangle\langle\psi|)$.*

Proof. Let Π be a projector with $|\mathcal{S}| = k$, and define $\mathbf{c} \in \{0,1\}^{n-1}$ such that $\mathbf{c}(j) = 1$ iff Π crosses the Schmidt cut at qudit j. For example, if Π acts on qudits $\{1,2\}$, then $\mathbf{c} = (1, 0, \ldots, 0)$. Note that in general $\| \mathbf{c} \|_1 = k - 1$. Let $|\psi_k\rangle$ denote the expression obtained by taking the RSD of $|\psi\rangle$ up to the kth level of recursion for $1 \leq k \leq n-1$, i.e. $|\psi_k\rangle$ can be written

$$|\psi_k\rangle = \sum_{i=1}^{d^k} \alpha_i \left|\psi_i^1\right\rangle \otimes \cdots \otimes \left|\psi_i^k\right\rangle \otimes |\phi_i\rangle, \tag{2.8}$$

where $\left|\psi_i^j\right\rangle \in \mathbb{C}^d$ and $|\phi_i\rangle \in (\mathbb{C}^d)^{\otimes n-k}$. (We assume $n \geq 2$, as otherwise the claim is vacuously true.) Corresponding to $|\psi_k\rangle$, define

$$\rho^{(k)} := \sum_{i=1}^{d^k} \alpha_i^2 \left|\psi_i^1\right\rangle\!\left\langle\psi_i^1\right| \otimes \cdots \otimes \left|\psi_i^k\right\rangle\!\left\langle\psi_i^k\right| \otimes |\phi_i\rangle\langle\phi_i|. \tag{2.9}$$

Define $c_k := \sum_{i=1}^{k} \mathbf{c}(i)$. To prove our claim, we show by induction that for all $1 \leq k \leq n-1$, it holds that

$$\mathrm{Tr}(\Pi |\psi\rangle\langle\psi|) \leq d^{c_k} \mathrm{Tr}(\Pi \rho^{(k)}). \tag{2.10}$$

Note that the case $k = n-1$ is in particular the case we are interested in.

For the base case, let $k = 1$. Consider first the Schmidt cut of $|\psi\rangle$ at qudit 1, i.e. $|\psi\rangle = \sum_{i=1}^{d} \alpha_i |\psi_i^1\rangle|\phi_i\rangle$, for $|\psi_i^1\rangle \in \mathbb{C}^d$ and $|\phi_i\rangle \in (\mathbb{C}^d)^{\otimes n-1}$. Then, recalling that $\rho^{(1)} = \sum_{i=1}^{d} \alpha_i^2 |\psi_i^1\rangle\langle\psi_i^1| \otimes |\phi_i\rangle\langle\phi_i|$, we have by Lemma 2.7 that

$$\mathrm{Tr}(\Pi|\psi\rangle\langle\psi|) \leq d^{\mathbf{c}(1)}\mathrm{Tr}(\Pi\rho^{(1)}), \tag{2.11}$$

as desired.

For the inductive step, assume the inductive hypothesis holds for some $1 \leq k \leq n-2$. We prove the claim holds for $k+1$. Note that by Equation (2.10), which holds due to the induction hypothesis for our specific value of k, it suffices to show that

$$\mathrm{Tr}(\Pi\rho^{(k)}) \leq d^{\mathbf{c}(k+1)}\mathrm{Tr}(\Pi\rho^{(k+1)}), \tag{2.12}$$

since $d^{c_k + \mathbf{c}(k+1)} = d^{c_{k+1}}$. To show this holds, consider the ith term in Equation (2.9), $|\psi_i^1\rangle\langle\psi_i^1| \otimes \cdots \otimes |\psi_i^k\rangle\langle\psi_i^k| \otimes |\phi_i\rangle\langle\phi_i|$, for arbitrary $1 \leq i \leq d^k$. Observe this term satisfies the preconditions for Corollary 2.8 with $m = k$. Hence, via Corollary 2.8 there exists a state σ_i acting on qudits $\{k+1, \ldots, n\}$ such that

$$\mathrm{Tr}(\Pi|\psi_i^1\rangle\langle\psi_i^1| \otimes \cdots \otimes |\psi_i^k\rangle\langle\psi_i^k| \otimes |\phi_i\rangle\langle\phi_i|) \leq d^{\mathbf{c}(k+1)}\mathrm{Tr}(\Pi|\psi_i^1\rangle\langle\psi_i^1| \otimes \cdots \otimes |\psi_i^k\rangle\langle\psi_i^k| \otimes \sigma_i). \tag{2.13}$$

Moreover, since σ_i in Corollary 2.8 is obtained via the Mixing Lemma (Lemma 2.7), by linearity we can express $\rho^{(k+1)}$ as

$$\rho^{(k+1)} = \sum_{i=1}^{d^k} \alpha_i^2 |\psi_i^1\rangle\langle\psi_i^1| \otimes \cdots \otimes |\psi_i^k\rangle\langle\psi_i^k| \otimes \sigma_i. \tag{2.14}$$

We conclude by linearity that Equation (2.12) holds, completing the proof. □

With Theorem 2.9 in hand, we can now show Theorem 2.2, i.e. that product states achieve approximation ratio $1/d^{k-1}$.

Proof. (Theorem 2.2) Simply apply Theorem 2.9 to each projector in the spectral decompositions of each (positive semidefinite) H_i in our MAX-k-local Hamiltonian instance $H = \sum_i H_i$, and let $|\psi\rangle$ denote the optimal assignment for H. It is important to note that we can exploit Theorem 2.9 in this fashion due to the fact that the ρ constructed by Theorem 2.9 is *independent* of the projector Π — i.e. for any fixed $|\psi\rangle$ and k, the state ρ provides the same approximation ratio against *any* k-local projector Π encountered in the spectral decompositions of the H_i. Finally, note that one can find a *pure* product state achieving this approximation guarantee since ρ is a convex mixture of pure product states. □

Upper bound of $d^{-\lfloor \frac{k}{2} \rfloor}$ for product state approximations. Is the result of Theorem 2.2 tight? In the case of MAX-2-local Hamiltonian on qudits, yes — consider a single clause projecting onto the maximally entangled state $\frac{1}{\sqrt{d}} \sum_i |ii\rangle$, for which a product state achieves value at most $1/d$. On the other hand, for MAX-3-local Hamiltonian on qubits, the worst case clause for a 3-qubit product state assignment is the projector onto the state $|W\rangle = \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle)$ [233]. But here product states achieve value $4/9$ [252], implying the bound of $1/4$ from Theorem 2.2 is not tight.

An upper bound on the true optimal ratio of $8k^2/(2^k)$ is implied by Theorem 2 of [120] for the case where $d = 2$ and $k \geq 11$. For general d and k, a simple construction shows that the optimal ratio is upper bounded by $d^{-\lfloor \frac{k}{2} \rfloor}$. To see this, consider a single clause which is the tensor product of maximally entangled bipartite states (for odd k, we assume the odd qudit out projects onto the identity). For example, for $n = 4$, consider the clause $|\phi^+\rangle\langle\phi^+| \otimes |\phi^+\rangle\langle\phi^+|$, where $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. The maximum value a product state can attain is $1/4$, as claimed. In the qubit setting ($d = 2$), one can further improve this construction for odd k by replacing the term $|\phi^+\rangle\langle\phi^+| \otimes I$ on the last three qubits with $|W\rangle\langle W|$. For example, for $k = 5$, setting our instance to be the clause $|\phi^+\rangle\langle\phi^+| \otimes |W\rangle\langle W|$ yields an upper bound of $(1/2)(4/9) = 2/9 < 1/4 = d^{-\lfloor \frac{k}{2} \rfloor}$ (where we again use the value $4/9$ for $|W\rangle$ from the previous paragraph). For general odd $k > 1$, this improved bound generalizes to $2^{\frac{-k+7}{2}}/9$.

2.3 Optimizing over the set of separable states

Section 2.2 showed that there always *exists* a product state assignment achieving a certain non-trivial approximation ratio. In this section, we show how to efficiently *find* such a product state. Our main theorem of this section is the following (Theorem 2.10), from which Theorem 2.4 follows easily (see discussion at end of Section 2.3.3). As the proofs and full notation of this section are rather dense, we first discuss our results below using simplified notation and without proofs. Full proofs and technical details are deferred to Section 2.4.

Theorem 2.10. *Let H be an instance of MAX-k-local Hamiltonian acting on n qudits, and let OPT_P denote the optimum value of $\mathrm{Tr}(H\rho)$ over all* product *states $\rho \in \mathcal{D}((\mathbb{C}^d)^{\otimes n})$. Then, for any fixed $\epsilon > 0$, there exists a polynomial time (deterministic) algorithm which outputs $\rho_1 \otimes \cdots \otimes \rho_n \in \mathcal{D}((\mathbb{C}^d)^{\otimes n})$ such that $\mathrm{Tr}(H\rho_1 \otimes \cdots \otimes \rho_n) \geq \mathrm{OPT}_P - \epsilon n^k$.*

We first outline our approach by generalizing the discussion in Section 2.1, introducing tools and notation we will require along the way. The optimal value OPT_P over prod-

uct state assignments for any MAX-k-local Hamiltonian instance can be expressed as the following program, denoted P_1:

$$\text{OPT}_P \;=\; \max \sum_{i_1,\ldots,i_k}^{n} \text{Tr}(H_{i_1,\ldots,i_k}\rho_{i_1}\otimes\cdots\otimes\rho_{i_k}) \text{ s.t. } \rho_i \succeq 0 \text{ and } \text{Tr}(\rho_i) = 1 \text{ for } 1 \leq i \leq n. \tag{2.15}$$

As done in Equation (2.3), we now recursively decompose our objective function as a sequence of nested sums. Let $\{\sigma_i\}_{i=1}^{d^2}$ be a Hermitian orthogonal basis for the set of Hermitian operators acting on $\mathbb{C}^d$, such that $\text{Tr}(\sigma_i\sigma_j) = 2\delta_{ij}$ (for δ_{ij} the Kroenecker delta). (See, e.g. [169], or Equations (6.5), (6.6), and (6.7) for an explicit construction of such basis elements. We remark that there is nothing special about the normalization factor of 2 in the term $2\delta_{ij}$ above; this value is simply consistent with the specific basis construction we have chosen to employ, which generalizes the Pauli basis for a qubit system.) Then, by rewriting each $H_{i_1,\ldots,i_k}$ in terms of $\{\sigma_i\}_{i=1}^{d^2}$, our objective function becomes

$$\sum_{i_k,\ldots,i_1}^{n} \text{Tr}\left[\left(\sum_{j_k,\ldots,j_1=1}^{d^2} r_{j_1,\ldots,j_k}^{i_1,\ldots,i_k}\sigma_{j_k}\otimes\cdots\otimes\sigma_{j_1}\right)\rho_{i_k}\otimes\cdots\otimes\rho_{i_1}\right] =$$
$$\sum_{i_k,j_k}\text{Tr}(\sigma_{j_k}\rho_{i_k})\left[\sum_{i_{k-1},j_{k-1}}\text{Tr}(\sigma_{j_{k-1}}\rho_{i_{k-1}})\left[\cdots\left[\sum_{i_1}\text{Tr}\left(\left(\sum_{j_1} r_{j_1,\ldots,j_k}^{i_1,\ldots,i_k}\sigma_{j_1}\right)\rho_{i_1}\right)\right]\right]\right], \tag{2.16}$$

where each $\mathbf{r}^{i_1,\ldots,i_k} \in \mathbb{R}^{d^2}$. We henceforth think of the objective function above as a "degree-k inner product", i.e. as a sequence of k nested sums involving inner products, in analogy to the degree-k polynomials of Reference [28]. In this sense, a degree-1 inner product would refer to only the innermost sums over i_1 and j_1, and a degree-k inner product would denote the entire expression in Equation (2.16). More formally, we denote a degree-b inner product for $1 \leq b \leq k$ using map $t_b : \mathcal{H}(\mathbb{C}^d)^{\times n} \mapsto \mathbb{R}$, defined such that

$$t_b(\rho_1,\ldots,\rho_n) := \sum_{i_b,j_b}\text{Tr}(\sigma_{j_b}\rho_{i_b})\left[\cdots\left[\sum_{i_1}\text{Tr}\left(\left(\sum_{j_1} r_{j_1,\ldots,j_k}^{i_1,\ldots,i_k}\sigma_{j_1}\right)\rho_{i_1}\right)\right]\right]. \tag{2.17}$$

Note that t_b implicitly depends on parameters $i_{b+1},\ldots,i_k$ and $j_{b+1},\ldots,j_k$. (See the beginning of Section 2.4 for more elaborate notation used in the proofs of the claims of Section 2.3.)

Our approach is to "linearize" the objective function of P_1 using exhaustive sampling and recursion to estimate its degree-$(k-1)$ inner products. To do so, we require the Sampling Lemma.

Lemma 2.11 (Sampling Lemma [28])**.** *Let (a_i) be a sequence of n real numbers with $|a_i| \leq M$ for all i, and let $f, g > 0$. If we choose a multiset of $s = g \log n$ of the a_i at random (with replacement), then their sum q satisfies $\sum_i a_i - nM\sqrt{\frac{f}{g}} \leq q \times \frac{n}{s} \leq \sum_i a_i + nM\sqrt{\frac{f}{g}}$ with probability at least $1 - n^{-f}$.*

The proof of Lemma 2.11 follows from a simple application of the Höffding bound [142]. To use the Sampling Lemma in conjunction with exhaustive sampling, we discretize the space of 1-qudit density operators using a δ-net $G \subseteq \mathcal{H}(\mathbb{C}^d)$, such that for all $\rho \in \mathcal{D}(\mathbb{C}^d)$, there exists $\sigma \in G$ such that $\| \rho - \sigma \|_{\mathrm{F}} \leq \delta$. We now show how to construct G.

To obtain G, we instead construct a δ-net for a subset of $\mathcal{H}(\mathbb{C}^d)$ which *contains* $\mathcal{D}(\mathbb{C}^d)$, namely the set $\mathcal{A}(\mathbb{C}^d) := \left\{A \in \mathcal{H}(\mathbb{C}^d) \mid \max_{i,j} |A(i,j)| \leq 1\right\}$. (Note: A net over $\mathcal{A}(\mathbb{C}^d)$ may allow non-positive assignments for a qudit. See Section 2.3.3 for why this is of no consequence.) Creating a δ-net over $\mathcal{A}(\mathbb{C}^d)$ is simple: we cast a (δ/d)-net over the unit disk for each of the complex $d(d-1)/2$ matrix entries above the diagonal, and likewise over $[-1, 1]$ for the entries on the diagonal. Letting m and n denote the minimum number of points required to create such (δ/d)-nets for each of the diagonal and off-diagonal entries, respectively, we have that $|G| = m^{\frac{d(d-1)}{2}} n^d$. For example, simple nets of size $m \approx d/\delta$ and $n \approx d^2/\delta^2$ can be obtained by placing a 1D and 2D grid over $[-1, 1]$ and the length 2 square in the complex plane centered at $(0, 0)$, respectively, implying $|G| \in O(1)$ when $d \in O(1)$. To show that G is indeed a δ-net, we now bound the Frobenius distance between arbitrary $\rho \in \mathcal{D}(\mathbb{C}^d)$ and the closest $\tilde{\rho} \in G$. (We use the Frobenius norm as it allows a simple analysis. Below, one could also consider the l_∞ norm bound $\| A \|_\infty \leq \delta/d$, where in this context $\| A \|_\infty = \max_{ij} |A(i,j)|$). Specifically, let $A := \rho - \tilde{\rho}$. Then:

$$\| A \|_{\mathrm{F}} = \sqrt{\mathrm{Tr}(A^\dagger A)} = \sqrt{\sum_{ij} |A(i,j)|^2} \leq \sqrt{\sum_{ij} (\delta/d)^2} = \frac{\delta}{d}(d) = \delta. \tag{2.18}$$

Finally, we remark that our *dense* assumption on MAX-k-local Hamiltonian instances is only necessary to convert the absolute error of Theorem 2.10 to a relative one (this conversion is detailed in Section 2.3.3). A dense assumption is not needed to apply the Sampling Lemma: Specifically, observe that Lemma 2.11 assumes there are n terms in the sum to be estimated, and that we are able to determine s of them. Looking back at Equation (2.1) and considering, say, qudit i, if we wish to use the Sampling Lemma to estimate the inner sum over neighbours $N(i)$ of i, we might run into a problem if i does *not* have $\Theta(n)$ neighbours. To circumvent this [28], observe that Lemma 2.11 only gives us an estimate to within $\pm\epsilon n$. Thus, if $N(i) \leq \epsilon n/10$ (say), then we do not use

the Sampling Lemma, but rather let our estimate be simply 0, which is guaranteed to fall within the desired error bounds (observe an estimate of 0 does not necessarily work, on the other hand, if $N(i)$ is large (say $N(i) = n - 1$), since typically $f/g < 1$). Throughout the remainder of our discussion, we assume this cutoff principle is implicitly present when employing Lemma 2.11.

The remaining sections of this chapter are organized as follows: In Section 2.3.1, we show how to recursively estimate degree-b inner products using the Sampling Lemma. We then use this estimation technique in Section 2.3.2 to linearize our optimization problem P_1. Section 2.3.3 brings everything together by presenting and analyzing the complete approximation algorithm. All technical proofs are found in Section 2.4.

2.3.1 Estimating degree-b inner products via sampling

Our recursive procedure, EVAL, for estimating a degree-b inner product using the Sampling Lemma is stated as Algorithm 2.12. There are two sources of error we must analyze: the Sampling Lemma, and our δ-net over $\mathbb{C}^d$. We claim that EVAL estimates the degree-b inner product $t_b(\rho_1, \ldots, \rho_n)$ to within additive error $\pm\epsilon_b n^b$, where ϵ_b is defined as follows. Set $\Delta := \sqrt{2}d(1+\delta)$, for δ from our δ-net. Then,

$$\epsilon_b := d^{\frac{k}{2}} \left(\sqrt{\frac{f}{g}} + \delta \right) \left(\frac{\Delta^b - 1}{\Delta - 1} \right). \tag{2.19}$$

The following lemma formalizes this claim. We adopt the convention of [28] and let $x \in y \pm z$ denote $x \in [y - z, y + z]$. Algorithm 2.12 is our operator analogue of the algorithm *Eval* in Section 3.3 of [28].

Lemma 2.13. *Let $t_k : \mathcal{H}(\mathbb{C}^k)^{\times n} \mapsto \mathbb{R}$ be defined using set $\{H_{i_1,\ldots,i_k}\} \subseteq \mathcal{H}((\mathbb{C}^d)^{\otimes k})$ (as in Equation (2.16)). Let $S \subseteq \{1, \ldots, n\}$ such that $|S| = g \log n$ have its elements chosen uniformly at random with replacement. Let $\rho_1, \ldots, \rho_n \in \mathcal{D}(\mathbb{C}^d)$ be some assignment on all n qudits, and $\{\tilde{\rho}_i : i \in S\}$ a set of elements in our δ-net such that $\| \rho_i - \tilde{\rho}_i \|_{\mathrm{F}} \leq \delta$ for all $i \in S$. Then, for $1 \leq b \leq k$, with probability at least $1 - d^{2b} n^{b-f}$, we have $\mathrm{EVAL}(t_b, S, \{\tilde{\rho}_i : i \in S\}) \in t_b(\rho_1, \ldots, \rho_n) \pm \epsilon_b n^b$, where ϵ_b is defined as in Equation (2.19).*

2.3.2 Linearizing our optimization problem

Our procedure, LINEARIZE, for "linearizing" the objective function of P_1 using EVAL from Section 2.3.1 is stated as Algorithm 2.14. Algorithm 2.14 takes as input P_1 and

Algorithm 2.12. EVAL(t_b , S , $\{\tilde{\rho}_i : i \in S\}$).

- Input: (1) A degree-b inner product $t_b : \mathcal{H}(\mathbb{C}^d)^{\times n} \mapsto \mathbb{R}$ for $1 \leq b \leq k$
  (2) A subset $S \subseteq \{1, \ldots, n\}$ of size $|S| = O(\log n)$
  (3) Sample points $\{\tilde{\rho}_i : i \in S\}$ such that $\| \tilde{\rho}_i - \rho_i \|_{\mathrm{F}} \leq \delta$ for all $i \in S$
- Output: $x \in \mathbb{R}$ such that $x \in t_b(\rho_1, \ldots, \rho_n) \pm \epsilon_b n^b$ (for ϵ_b defined in Equation (2.19)).

1. (Base Case) If $b = 1$, return $\frac{n}{|S|} \sum_{i_1 \in S} \mathrm{Tr}\left(\left(\sum_{j_1=1}^{d^2} r^{i_1,\ldots,i_k}_{j_1,\ldots,j_k} \sigma_{j_1}\right) \rho_{i_1}\right)$. (Note this return value depends on $i_2, \ldots, i_k, j_2, \ldots, j_k$, which are assumed to have a fixed value in the current recursive call to EVAL.)
2. (Recurse) For all $i \in S$ and $j = 1 \ldots d^2$, set $e_{ij} = \mathrm{EVAL}(t^{ij}_{b-1}, S, \{\tilde{\rho}_i : i \in S\})$, where t^{ij}_{b-1} is the term to the right of $\mathrm{Tr}(\sigma_{j_b} \rho_{i_b})$ in Equation (2.17).
3. Return $\frac{n}{|S|} \sum_{i \in S} \left[\sum_{j=1}^{d^2} \mathrm{Tr}(\sigma_j \tilde{\rho}_i) e_{ij}\right]$.

a set of sample points $\{\tilde{\rho}_i\}$, and outputs a semidefinite program (SDP) which we shall henceforth refer to as P_2. We remark that LINEARIZE is our version of the procedure *Linearize* in Section 3.4 of [28], extended to the setting of operators and a more complex error structure. Although LINEARIZE is presented as linearizing an objective function here, the same techniques straightforwardly apply in linearizing constraints involving high-degree inner products.

We remark that the linear constraints output on each recursive call on line 3(b) of Algorithm 2.14 ensure the approximate consistency with our estimates from EVAL for any solution to P_2, as well as play a crucial role in bounding how good of an approximation P_2 yields to P_1.

To prove correctness of our final approximation algorithm, we require the following two important lemmas regarding P_2. The first shows that any feasible solution $(\rho_1, \ldots, \rho_n)$ for P_1 consistent with the sample set $\{\tilde{\rho}_i : i \in S\}$ fed into LINEARIZE is also a feasible solution for P_2 with high probability.

Lemma 2.15. *Let t_k, assignment $(\rho_1, \ldots, \rho_n)$, S, and $\{\tilde{\rho}_i : i \in S\}$ be defined as in Lemma 2.13. Then, for any $f, g > 0$, calling LINEARIZE with parameters t_k, $\{\tilde{\rho}_i : i \in S\}$, and $\epsilon = \epsilon_k$ (for ϵ_k defined in Equation (2.19)) yields an SDP P_2 for which the assignment $\{\rho_1, \ldots, \rho_n\}$ is feasible with probability at least $1 - d^{2k} n^{k-f}$.*

Algorithm 2.14. LINEARIZE(t_b , $\mathcal{N}$, S, $\{\tilde{\rho}_i : i \in S\}$, ϵ, U, L).

- Input: (1) A degree-b inner product $t_b : \mathcal{H}(\mathbb{C}^d)^{\times n} \mapsto \mathbb{R}$ for $1 \leq b \leq k$.
  (2) A set of linear constraints $\mathcal{N}$ (e.g. "$\rho_i \succeq 0$").
  (3) A subset $S \subseteq \{1, \ldots, n\}$ of size $|S| = O(\log n)$.
  (4) Sample points $\{\tilde{\rho}_i : i \in S\}$ consistent with some feasible solution $(\rho_1, \ldots, \rho_n)$ for P_1 such that $\| \tilde{\rho}_i - \rho_i \|_{\mathrm{F}} \leq \delta$ for all $i \in S$.
  (5) An error parameter $\epsilon > 0$.
  (6) (Optional) upper and lower bounds $U, L \in \mathbb{R}$. If U and L are not provided, we assume $U, L = \infty$.

- Output: (1) (Optional) A linear objective function $f : (\mathcal{L}(\mathbb{C}^d))^{\times n} \to \mathbb{R}$.
  (2) An updated set of linear constraints, $\mathcal{N}$.

1. (Base case) If $b = 1$, then
   (a) (Trivial: Initial objective function was linear) If $U = L = \infty$, return $[t_b, \mathcal{N}]$.
   (b) (Reached bottom of recursion) Else, return $[\mathcal{N} \cup \{\text{“}L \leq t_b(\rho_1, \ldots, \rho_n) \leq U\text{”}\}]$.
2. (Recursive case) For $i = 1 \ldots n$ and $j = 1 \ldots d^2$ do
   (a) Set $e_{ij} := \mathrm{EVAL}(t^{ij}_{b-1}, S, \{\tilde{\rho}_i : i \in S\})$.
   (b) Set $\epsilon' := \epsilon - d^{\frac{k}{2}} \left(\sqrt{\frac{f}{g}} + \delta \right) \Delta^{b-1}$, for Δ defined in Equation (2.19).
   (c) Set $l_{ij} := e_{ij} - \epsilon' n^{b-1}$ and $u_{ij} := e_{ij} + \epsilon' n^{b-1}$.
   (d) Call LINEARIZE$(t^{ij}_{b-1}, \mathcal{N}, S, \{\tilde{\rho}_i : i \in S\}, \epsilon', u_{ij}, l_{ij})$.
3. (a) (Entire computation done) If $U = L = \infty$, return $\left[\sum_{ij} \mathrm{Tr}(\sigma_j \rho_i) e_{ij}, \mathcal{N} \right]$.
   (b) (Recursive call done) Else, return
   $$\left[\mathcal{N} \cup \left\{ \text{“}L - \epsilon' d^2 n^b \leq \textstyle\sum_{ij} \mathrm{Tr}(\sigma_j \rho_i) e_{ij} \leq U + \epsilon' d^2 n^b\text{”} \right\} \right].$$

The second lemma is a bound on how far the optimal solution of P_2 is from the optimal solution for P_1. We adopt the convention of [28] and write $[x, y] \pm z$ to denote interval $[x - z, y + z]$.

Lemma 2.16. *Let* OPT_P *be the optimal value for* P_1*, obtained by assignment* $\rho^{\mathrm{OPT}_P} := (\rho_1^{\mathrm{opt}}, \ldots, \rho_n^{\mathrm{opt}})$*. Let assignment* $\{\rho_i\}_{i=1}^n = \{\rho_i^{\mathrm{opt}}\}_{i=1}^n$*,* S*, and* $\{\tilde{\rho}_i : i \in S\}$ *be defined as in Lemma 2.13. Let* P_2 *denote the SDP obtained by calling LINEARIZE with* S*, and denote by* ϵ_m *for* $1 \leq m \leq k$ *the error parameter passed with map* t_m *into a (possibly recursive) call to LINEARIZE. Then, letting* OPT_2 *denote the optimal value of* P_2*, we*

Algorithm 2.17. APPROXIMATE(H , ϵ).

- Input: (1) A k-local Hamiltonian $H = \sum_{i_1,\ldots,i_k} H_{i_1,\ldots,i_k}$ for each $H_{i_1,\ldots,i_k} \in \mathcal{H}((\mathbb{C}^d)^{\otimes k})$.
  (2) An error parameter $\epsilon > 0$.
- Output: A product assignment $\rho_1 \otimes \cdots \otimes \rho_n$ that with probability at least $1/2$, has value at least $\mathrm{OPT}_P - \epsilon n^k$, for OPT_P the optimal value for H over all product state assignments.

1. Set $\varepsilon_{\mathrm{sdp}} := \varepsilon/10$.
2. Define $h : \mathbb{R} \to \mathbb{R}$ such that for any error parameter ϵ input to LINEARIZE, $h(\varepsilon)n^k$ is the absolute value of the bound on additive error given by Lemma 2.16. Then, define ϵ' implicitly so that $h(\epsilon') + \varepsilon_{\mathrm{sdp}} = \epsilon$ holds.
3. Define constant f such that $1 - d^{2k}n^{k-f} > 1/2$.
4. Define constants g and δ implicitly so that $\epsilon' = d^{\frac{k}{2}} \left(\sqrt{\frac{f}{g}} + \delta\right) \left(\frac{\Delta^k - 1}{\Delta - 1}\right)$, for Δ defined in Equation (2.19).
5. Choose $g \log n$ indices $S \subseteq \{1, \ldots, n\}$ independently and uniformly at random.
6. For each possible assignment i from our δ-net to the qudits in S:
   (a) Call LINEARIZE$(t_k, \{P_1\text{'s constraints}\}, S, i, \epsilon')$ to obtain SDP P_2^i.
   (b) Let α_i denote the value of P_1 obtained by substituting in the optimal solution of P_2^i.
7. Return the assignment corresponding to the maximum over all α_i.

have with probability at least $1 - d^{2k}n^{k-f}$ *(for parameters set as in Lemma 2.15) that* $\mathrm{OPT}_2 \in \mathrm{OPT}_P \pm d(d + \sqrt{2}) \left[\sum_{m=1}^{k-1} (\sqrt{2}d)^{k-1-m} \epsilon_m\right] n^k$.

2.3.3 The final algorithm

We finally present our approximation algorithm, APPROXIMATE (Algorithm 2.17), in its entirety, which exploits our ability to linearize P_1 using LINEARIZE (Algorithm 2.14). This proves Theorem 2.10, which in turn implies Theorem 2.4. We first clarify a few points about APPROXIMATE, then analyze its runtime, and follow with further discussion, including the algorithm's derandomization and a proof that dense MAX-k-local Hamiltonian remains QMA-hard.

We begin by explaining the rationale behind the constants in Algorithm 2.17. The constant ε_{sdp} is the additive error incurred when solving an SDP [121]. We choose ϵ' so that after running LINEARIZE and solving P_2^i, the total additive error is at most ϵ, as desired. We choose f to ensure the probability of success is at least $1/2$. Finally, we set g large enough and δ (for our δ-net) small enough to ensure that ϵ' matches the error bounds for EVAL in Lemma 2.13.

We now analyze the runtime of Algorithm 2.17. Let $|G|$ denote the size of our δ-net G for a qudit. Then, for each of the $|G|^{g \log n}$ iterations of line 6, we first take $O(n^{k-1})$ time to run LINEARIZE, outputting $O(n^{k-1})$ new linear constraints (seen via a simple inductive argument). We then solve SDP P_2^i, which can be done in time polynomial in n and $\log(1/\varepsilon_{\text{sdp}})$ using the ellipsoid method [121] (see, e.g., [249]). Let $r(n, \varepsilon_{\text{sdp}})$ denote the maximum runtime required to solve any of the P_2^i. Then, the overall runtime for Algorithm 2.17 is $O(n^{g \log |G|}(n^{k-1} + r(n, \varepsilon_{\text{sdp}})))$, which is polynomial in n for $\epsilon, d, k \in O(1)$ (recall from Section 2.3 that $|G| \in O((\frac{d}{\delta})^d)$, and that δ and g are constant in our setting). Note that, due to the implicit dependence of g on ϵ, this runtime scales at least exponentially with varying ϵ.

Before moving to further discussion, we make two remarks. First, one can efficiently convert the output of Algorithm 2.17 to a *pure* state with the same guarantee by adapting the standard classical *method of conditional expectations* [236]. To demonstrate, suppose $\{\rho_i\}$ is output by Algorithm 2.17. Then, set ρ_1' to be the eigenvector $|\psi_j\rangle\langle\psi_j|$ of ρ_1 for which the assignment $|\psi_j\rangle\langle\psi_j| \otimes \rho_2 \otimes \cdots \otimes \rho_n$ performs best for P_1. (If the spectrum of ρ_i is degenerate, begin by fixing an arbitrary choice of spectral decomposition for ρ_i.) Let our new assignment be $\rho_1' \otimes \rho_2 \otimes \cdots \otimes \rho_n$. Now repeat for each ρ_i for $2 \leq i \leq n$. The final state $\rho_1' \otimes \cdots \otimes \rho_n'$ is pure, and by convexity is guaranteed to perform as well as $\rho_1 \otimes \cdots \otimes \rho_n$.

Second, recall from Section 2.3 that we constructed a δ-net over a space larger than $\mathcal{D}(\mathbb{C}^d)$, allowing possibly non-positive assignments for a qudit. We now see that this is of no consequence, since regardless of which samples (positive or not) we use to derive our estimates with the Sampling Lemma, any feasible solution to P_2^i in Algorithm 2.17 is a valid assignment for P_1. Moreover, we know that for each optimal ρ_i for P_1, there must be *some* operator (positive or not) within distance δ in our net, ensuring our estimates obtained using the Sampling Lemma are within our error bounds.

Converting the absolute error of Algorithm 2.17 into relative error. To convert the absolute error $\pm\epsilon n^k$ of Algorithm 2.17 into a *relative* error of $1 - \epsilon'$ for any ϵ', define constant c such that cn^k is the value obtained for a MAX-k-local Hamiltonian instance by choosing the maximally mixed assignment I/d^n (analogous to a classical random as-

signment). Since I/d^n can be written as a mixture of computational basis states, we have $\mathrm{OPT}_P \geq cn^k$. It follows that by setting $\epsilon = c\epsilon'$, Algorithm 2.17 returns an assignment with value at least $\mathrm{OPT}_P - c\epsilon' n^k \geq \mathrm{OPT}_P - \epsilon' \mathrm{OPT}_P \geq \mathrm{OPT}_P(1-\epsilon')$, as desired.

Derandomizing Algorithm 2.17. The source of randomness in our algorithm is Lemma 2.11. By a standard argument in [28] (see also [42, 41]), this randomness can be eliminated with only polynomial overhead. Specifically, we replace the random selection of $g \log n$ indices in the Sampling Lemma with the set of indices encountered on a random walk of length $O(g \log n)$ along a constant degree expander [113]. Since the expander has constant degree, we can efficiently deterministically iterate through all $n^{O(g)}$ such walks, and since such a walk works with probability $1/n^{O(1)}$, at least one walk will work for all $\mathrm{poly}(n)$ sampling experiments we wish to run.

QMA-hardness of dense MAX-k-local Hamiltonian. It is easy to see that (exact) MAX-2-local Hamiltonian remains QMA-hard for dense instances (a similar statement holds for MAX-2-SAT [28]). For any MAX-2-local Hamiltonian instance with optimal value OPT, we simply add n qudits, between any two of which we place the constraint $|00\rangle\langle 00|$ (no constraints are necessary between old and new qudits). Then, the new Hamiltonian has optimal value $\mathrm{OPT} + \binom{n}{2}$, making it dense, and the ability to solve this new instance implies the ability to solve the original one. The argument extends straightforwardly to MAX-k-local Hamiltonian for $k > 2$.

2.4 Further technical details and proofs

We now prove our claims in Section 2.3. For this, we first require expanding on the notation we have set thus far.

Expanded Notation. We now expand on our previous notation for analyzing Equation (2.16) in order to facilitate proofs of the claims in Section 2.3. First, to recursively analyze a clause $H_{i_1,\ldots,i_k} \subseteq \mathcal{H}((\mathbb{C}^d)^{\otimes k})$, let $H_b \in \mathcal{H}((\mathbb{C}^d)^{\otimes b})$ for any $1 \leq b \leq k$ denote the action of $H_{i_1,\ldots,i_k}$ restricted to the first b of its k target qudits, i.e.

$$H_b := \sum_{j_b,\ldots,j_1=1}^{d^2} r^{i_1,\ldots,i_k}_{j_1,\ldots,j_k} \sigma_{j_b} \otimes \cdots \otimes \sigma_{j_1}. \tag{2.20}$$

For example, $H_1 = \sum_{j_1=1}^{d^2} r_{j_1,\ldots,j_k}^{i_1,\ldots,i_k} \sigma_{j_1}$ and $H_k = H_{i_1,\ldots,i_k}$. Note that H_b implicitly depends on variables $i_1, \ldots, i_k, j_{b+1}, \ldots, j_k$. To reduce clutter, however, our notation does not explicitly denote this dependence unless necessary. Next, to recursively analyze a degree-a inner product, we define $t_{a,b} : \mathcal{H}(\mathbb{C}^d)^{\times n} \mapsto \mathbb{R}$ for any $0 \leq a \leq k$ and $1 \leq b \leq k$ such that

$$t_{a,b}(\rho_1, \ldots, \rho_n) := \sum_{i_a,\ldots,i_1=1}^{n} \mathrm{Tr}\left(H_b^{i_1,\ldots,i_k} \rho_{i_b} \otimes \cdots \otimes \rho_{i_1}\right) \tag{2.21}$$

(where setting $a = 0$ eliminates the sum over indices i). For example, $t_{k,k}$ is our full "degree-k" objective function in Equation (2.15), and more generally, $t_{b,b}$ is the degree-b inner product in Equation (2.16). Allowing different values for a and b greatly eases our technical analysis. We use the shorthand t_b to denote $t_{b,b}$, and again only explicitly denote the dependence of $t_{a,b}$ on parameters $i_{a+1}, \ldots, i_k$ and $j_{b+1}, \ldots, j_k$ when necessary.

We now state and prove a technical lemma required for the remainder of our proofs here.

Lemma 2.18. *Let $\{\rho_i\}_{i=1}^n \subseteq \mathcal{H}(\mathbb{C}^d)$. For $\{H_{i_1,\ldots,i_k}\} \subseteq \mathcal{H}(\mathbb{C}^{d^k})$ any MAX-k-local Hamiltonian instance with decomposition for the $H_{i_1,\ldots,i_k}$ as given in Equation (2.16), we have for any $0 \leq a \leq k$ and $1 \leq b \leq k$ that $|t_{a,b}(\rho_1, \ldots, \rho_n)| \leq \left(\max_{i_b,\ldots,i_1} \| \rho_{i_b} \|_{\mathrm{F}} \cdots \| \rho_{i_1} \|_{\mathrm{F}}\right) d^{\frac{k}{2}} n^a$.*

Proof. By the triangle inequality and the Hölder inequality for Schatten p-norms (see Section 1.3), we have

$$\begin{aligned} |t_{a,b}| = \left| \sum_{i_a,\ldots,i_1=1}^{n} \mathrm{Tr}\left(H_b \rho_{i_b} \otimes \cdots \otimes \rho_{i_1}\right) \right| &\leq \sum_{i_a,\ldots,i_1=1}^{n} \| H_b \|_{\mathrm{F}} \| \rho_{i_b} \otimes \cdots \otimes \rho_{i_1} \|_{\mathrm{F}} \\ &\leq \left(\max_{i_b,\ldots,i_1} \| \rho_{i_b} \|_{\mathrm{F}} \cdots \| \rho_{i_1} \|_{\mathrm{F}} \right) \sum_{i_a,\ldots,i_1=1}^{n} \| H_b \|_{\mathrm{F}}, \end{aligned} \tag{2.22}$$

where we have used the fact that $\| A \otimes B \|_{\mathrm{F}} = \| A \|_{\mathrm{F}} \| B \|_{\mathrm{F}}$ for all $A, B \in \mathcal{L}(\mathbb{C}^d)$. If we can now show that $\| H_b \|_{\mathrm{F}} \leq \| H_k \|_{\mathrm{F}}$ for all $1 \leq b \leq k$, then we would be done since we would have $\sum_{i_a,\ldots,i_1}^{n} \| H_b \|_{\mathrm{F}} \leq \| H_k \|_{\mathrm{F}} n^a \leq d^{\frac{k}{2}} n^a$, where $\| H_k \|_{\mathrm{F}} \leq d^{\frac{k}{2}}$ since $\| H_k \|_{\infty} \leq 1$ by definition. Indeed, we claim that for any fixed $1 \leq b \leq k$, we have $\| H_b \|_{\mathrm{F}} \leq 2^{\frac{b-k}{2}} \| H_k \|_{\mathrm{F}}$. To see this, note by straightforward expansion of the Frobenius norm and the fact that $\mathrm{Tr}(\sigma_i \sigma_j) = 2\delta_{ij}$ that

$$\| H_b \|_{\mathrm{F}} = \sqrt{\mathrm{Tr}(H_b^2)} = 2^{\frac{b}{2}} \sqrt{\sum_{j_b,\ldots,j_k} (r_{j_1,\ldots,j_k}^{i_1,\ldots,i_k})^2} \leq 2^{\frac{b}{2}} \left\| \mathbf{r}^{i_1,\ldots,i_k} \right\|_2 = 2^{\frac{b-k}{2}} \left(2^{\frac{k}{2}} \left\| \mathbf{r}^{i_1,\ldots,i_k} \right\|_2 \right), \tag{2.23}$$

where $\mathbf{r}^{i_1,\ldots,i_k}$ is the coordinate vector of $H_{i_1,\ldots,i_k}$ from Equation (2.16). Note, however, then for $b = k$, the inequality in the chain above is an equality, and so $\| H_k \|_{\mathrm{F}} = 2^{\frac{k}{2}} \| \mathbf{r}^{i_1,\ldots,i_k} \|_2$. Substituting this into the chain above completes the proof of our claim. □

We now prove our claims of Section 2.3.

Proof of Lemma 2.13. We first derive the error bound of ϵ_b, and subsequently prove the probability bound. We follow [28], and proceed by induction on b. For the base case $b = 1$, EVAL$(H_1, S, \{\tilde{\rho}_i : i \in S\})$ attempts to estimate

$$t_1(\rho_1, \ldots, \rho_n) = \sum_{i_1} \left[\sum_{j_1} r^{i_1,\ldots,i_k}_{j_1,\ldots,j_k} \mathrm{Tr}(\sigma_{j_1} \rho_{i_1}) \right] \tag{2.24}$$

using our flawed sample points $\{\tilde{\rho}_i : i \in S\}$. To analyze the error of its output, assume first that our sample points are exact, i.e. $\tilde{\rho}_i = \rho_i$ for all $i \in S$. Then, by setting "a_i" in Lemma 2.11 to $t^{i_1}_{0,1}$ for $i = i_1$, and by using Lemma 2.18 with parameters $a = 0$ and $b = 1$ to obtain upper bound $M = d^{\frac{k}{2}}$, we have by the Sampling Lemma that (with probability at least $1 - n^{-f}$)

$$\frac{n}{|S|} \sum_{i_1 \in S} \left[\sum_{j_1} r^{i_1,\ldots,i_k}_{j_1,\ldots,j_k} \mathrm{Tr}(\sigma_{j_1} \rho_{i_1}) \right] \in t_1(\rho_1, \ldots, \rho_n) \pm d^{\frac{k}{2}} \sqrt{\frac{f}{g}} n. \tag{2.25}$$

(Recall that the notation $x \in y \pm z$ means here $x \in [y - z, y + z]$.) This bound holds if we sum over exact sample points. If we instead sum over flawed sample points $\{\tilde{\rho}_i : i \in S\}$, the additional error is bounded by $\frac{n}{|S|}$ times

$$\begin{aligned}
\left| \sum_{i_1 \in S} \left[\sum_{j_1} r^{i_1,\ldots,i_k}_{j_1,\ldots,j_k} \mathrm{Tr}(\sigma_{j_1} (\rho_{i_1} - \tilde{\rho}_{i_1})) \right] \right| &\leq \sum_{i_1 \in S} \left| \sum_{j_1} r^{i_1,\ldots,i_k}_{j_1,\ldots,j_k} \mathrm{Tr}(\sigma_{j_1} (\rho_{i_1} - \tilde{\rho}_{i_1})) \right| && (2.26) \\
&\leq \sum_{i_1 \in S} (\| \rho_{i_1} - \tilde{\rho}_{i_1} \|_{\mathrm{F}} \, d^{\frac{k}{2}}) && (2.27) \\
&\leq d^{\frac{k}{2}} \delta n, && (2.28)
\end{aligned}$$

where the second inequality uses Lemma 2.18 with parameters $a = 0$ and $b = 1$ and the

promise of our δ-net. We conclude for the base case that, as desired,

$$\begin{aligned} \mathrm{EVAL}(H_1, S, \{\tilde{\rho}_i : i \in S\}) &= \frac{n}{|S|} \sum_{i_1 \in S} \left[\sum_{j_1} r^{i_1,\ldots,i_k}_{j_1,\ldots,j_k} \mathrm{Tr}(\sigma_{j_1} \tilde{\rho}_{i_1}) \right] && (2.29) \\ &\in t_1(\rho_1, \ldots, \rho_n) \pm d^{\frac{k}{2}} \left(\sqrt{\frac{f}{g}} + \delta \right) n. && (2.30) \end{aligned}$$

Assume now that the inductive hypothesis holds for $1 \leq m \leq b-1$. We prove the claim for $m = b$. To do so, suppose first that the recursive calls on line 1(b) of Algorithm 2.12 return the *exact* values of $t^{ij}_{b-1}(\rho_1, \ldots, \rho_n)$, and that we have exact samples $\{\rho_i : i \in S\}$. Then, since by calling Lemma 2.18 with $a = b-1$ we have $\left| \sum_j \mathrm{Tr}(\sigma_j \rho_i) t^{ij}_{b-1}(\rho_1, \ldots, \rho_n) \right| \leq d^{\frac{k}{2}} n^{b-1}$, it follows by the Sampling Lemma that

$$\frac{n}{|S|} \sum_{i \in S} \left[\sum_j \mathrm{Tr}(\sigma_j \rho_i) t^{ij}_{b-1}(\rho_1, \ldots, \rho_n) \right] \in \sum_{i=1}^{n} \left[\sum_j \mathrm{Tr}(\sigma_j \rho_i) t^{ij}_{b-1}(\rho_1, \ldots, \rho_n) \right] \pm d^{\frac{k}{2}} \sqrt{\frac{f}{g}} n^b. \tag{2.31}$$

To first adjust for using flawed samples, observe that an analogous calculation to Equation (2.28) yields $\left| \frac{n}{|S|} \sum_{i \in S} \left[\sum_j \mathrm{Tr}(\sigma_j (\rho_i - \tilde{\rho}_i)) \right] \right| \leq d^{\frac{k}{2}} \delta n^b$, where we have called Lemma 2.18 with $a = b - 1$. Thus, using flawed samples, the output of Algorithm 2.12 satisfies

$$\frac{n}{|S|} \sum_{i \in S} \left[\sum_j \mathrm{Tr}(\sigma_j \tilde{\rho}_i) t^{ij}_{b-1} \right] \in \sum_{i=1}^{n} \left[\sum_j \mathrm{Tr}(\sigma_j \rho_i) t^{ij}_{b-1} \right] \pm d^{\frac{k}{2}} \left(\sqrt{\frac{f}{g}} + \delta \right) n^b. \tag{2.32}$$

To next drop the assumption that our estimates e_{ij} on line 1(b) are exact, apply the induction hypothesis to conclude that $e_{ij} \in t^{ij}_{b-1}(\rho_1, \ldots, \rho_n) \pm \epsilon_{b-1} n^{b-1}$. Then,

$$\begin{aligned} \frac{n}{|S|} \sum_{i \in S} \left[\sum_j \mathrm{Tr}(\sigma_j \tilde{\rho}_i) e_{ij} \right] &\in \frac{n}{|S|} \sum_{i \in S} \left[\sum_j \mathrm{Tr}(\sigma_j \tilde{\rho}_{ij}) \left(t^{ij}_{b-1} \pm \epsilon_{b-1} n^{b-1} \right) \right] \\ &\subseteq \frac{n}{|S|} \sum_{i \in S} \left[\sum_j \mathrm{Tr}(\sigma_j \tilde{\rho}_i) t^{ij}_{b-1} \right] \pm \frac{\epsilon_{b-1} n^b}{|S|} \sum_{i \in S} \left[\sum_{j=1}^{d^2} \mathrm{Tr}(\sigma_j \tilde{\rho}_i) \right] \\ &\subseteq \frac{n}{|S|} \sum_{i \in S} \left[\sum_j \mathrm{Tr}(\sigma_j \tilde{\rho}_i) t^{ij}_{b-1} \right] \pm \epsilon_{b-1} \sqrt{2} d (1 + \delta) n^b, && (2.33) \end{aligned}$$

where the last statement follows since

$$\left|\sum_{j=1}^{d^2} \mathrm{Tr}(\sigma_j \tilde{\rho}_i)\right| = \left|\sum_{j=1}^{d^2} \mathrm{Tr}\left(\sigma_j \left(\sum_{m=1}^{d^2} \tilde{r}_m \sigma_m\right)\right)\right| \leq 2\sum_{m=1}^{d^2} |\tilde{r}_m| \leq 2d \,\|\, \tilde{\mathbf{r}} \,\|_2 \leq \sqrt{2}d(1+\delta), \tag{2.34}$$

where $\tilde{\mathbf{r}}$ denotes the coordinate vector of $\tilde{\rho}_i$ with respect to basis $\{\sigma_m\}$, and we have used the facts that $\mathrm{Tr}(\sigma_i \sigma_j) = 2\delta_{ij}$, that $\|\, \mathbf{x} \,\|_1 \leq \sqrt{d} \,\|\, \mathbf{x} \,\|_2$ for $\mathbf{x} \in \mathbb{C}^d$, that $\|\, \tilde{\rho}_i \,\|_{\mathrm{F}} = \sqrt{2} \,\|\, \tilde{\mathbf{r}} \,\|_2$ for any $\tilde{\rho}_i \in \mathcal{H}(\mathbb{C}^d)$, and that $\|\, \tilde{\rho}_i \,\|_{\mathrm{F}} \leq 1 + \delta$ (which follows from our δ-net and the triangle inequality). Thus, recalling that $\Delta = \sqrt{2}d(1+\delta)$ and substituting Equation (2.32) into Equation (2.33), we have that

$$\frac{n}{|S|} \sum_{i \in S} \left[\sum_j \mathrm{Tr}(\sigma_j \tilde{\rho}_i) e_{ij}\right] \in t_b(\rho_1, \ldots, \rho_n) \pm \left[d^{\frac{k}{2}} \left(\sqrt{\frac{f}{g}} + \delta\right) + \epsilon_{b-1}\Delta\right] n^b. \tag{2.35}$$

We hence have the recurrence relation $\epsilon_b \leq d^{\frac{k}{2}} \left(\sqrt{\frac{f}{g}} + \delta\right) + \epsilon_{b-1}\Delta$, which when unrolled yields

$$\epsilon_b \leq d^{\frac{k}{2}} \left(\sqrt{\frac{f}{g}} + \delta\right) \sum_{m=0}^{b-1} \Delta^m = d^{\frac{k}{2}} \left(\sqrt{\frac{f}{g}} + \delta\right) \left(\frac{\Delta^b - 1}{\Delta - 1}\right), \tag{2.36}$$

as desired. This concludes the proof of the error bound.

To prove the probability bound, we show a stronger bound of $1 - (\sum_{m=0}^{b-1} d^{2m} n^m) n^{-f}$ by induction on b. The base case $b = 1$ follows directly from our application of the Sampling Lemma in Equation (2.25). For the inductive step, define for brevity of notation $\gamma := d^2 n$, and apply the induction hypothesis to line 1(b) of Algorithm 2.12 to conclude that each of the γ calls to EVAL fails will probability at most $(\sum_{m=0}^{b-2} \gamma^m) n^{-f}$. Then, by the union bound, the probability that at least one call fails is at most $(\sum_{m=1}^{b-1} \gamma^m) n^{-f}$. Similarly, since our application of the Sampling Lemma in line 2 of Algorithm 2.12 fails with probability at most n^{-f}, we arrive at our claimed stronger bound of $1 - \left(\sum_{m=0}^{b-1} \gamma^m\right) n^{-f}$, as desired. □

Proof of Lemma 2.15. We begin by observing that if one sets $\epsilon = \epsilon_k$, then the value of ϵ' in line 2(b) of Algorithm 2.14 is precisely ϵ_{k-1}, and more generally, the ϵ passed into the recursive call of line 2(e) on t_b for any $1 \leq b \leq k$ is ϵ_b. Now, focus on some recursive call on t_b for $b > 1$ (the case of $b = 1$ is straightforward by Lemma 2.13). If the estimates e_{ij} in line 2(a) succeed, then by Lemma 2.13, we know that $e_{ij} \in t^{ij}_{b-1}(\rho_1, \ldots, \rho_n) \pm \epsilon_{b-1} n^{b-1}$, implying $t^{ij}_{b-1}(\rho_1, \ldots, \rho_n) \in [l_{ij}, u_{ij}]$. Now, l_{ij} and u_{ij} are only incorporated into linear constraints in

recursive calls on t_{b-1}^{ij}, yielding constraints of the form

$$l_{i_b j_b} - \epsilon_{b-2} d^2 n^{b-1} \leq \sum_{i_{b-1}, j_{b-1}} \mathrm{Tr}(\sigma_{j_{b-1}} \rho_{i_{b-1}}) e_{i_{b-1} j_{b-1}} \leq u_{i_b j_b} + \epsilon_{b-2} d^2 n^{b-1}. \tag{2.37}$$

But $\{\rho_1, \ldots, \rho_n\}$ must now satisfy this constraint, since recall

$$t_{b-1}(\rho_1, \ldots, \rho_n) = \sum_{i_{b-1}, j_{b-1}} \mathrm{Tr}(\sigma_{j_{b-1}} \rho_{i_{b-1}}) t_{b-2}^{i_{b-1} j_{b-1}}(\rho_1, \ldots, \rho_n), \tag{2.38}$$

and there are $d^2 n$ terms $e_{i_{b-1} j_{b-1}}$ in Equation (2.37) each yielding an additional error of at most $\epsilon_{b-2} n^{b-2}$ (assuming EVAL succeeded on $t_{b-2}^{i_{b-1} j_{b-1}}$ in line 2(a)) above and beyond the bounds $t_{b-1}^{ij}(\rho_1, \ldots, \rho_n) \in [l_{ij}, u_{ij}]$ we established above.

We conclude that if, for *all* b, i, and j, EVAL succeeds in producing estimates $e_{i_b}^{ij}$, then $\{\rho_1, \ldots, \rho_n\}$ is a feasible solution for P_2, as desired. The probability of this happening is, by the proof of Lemma 2.13, at least $1 - d^{2k} n^{k-f}$, since EVAL recursively estimates precisely the same terms during its execution[1]. □

Proof of Lemma 2.16. We begin by proving that for any recursive call to LINEARIZE on t_b with valid upper and lower bounds U and L (i.e. $U, L \neq \infty$), respectively, we have for *any* feasible solution $(\rho_1, \ldots, \rho_n)$ to P_2 that

$$t_b(\rho_1, \ldots, \rho_n) \in [L, U] \pm d(d + \sqrt{2}) \left[\sum_{m=1}^{b-1} (\sqrt{2} d)^{b-1-m} \epsilon_m \right] n^b. \tag{2.39}$$

We prove this by induction on b, following [28]. For base case $b = 1$, the claim is trivial by line 1(b) of the algorithm. Now, assume by induction hypothesis that

$$t_{b-1}^{ij}(\rho_1, \ldots, \rho_n) \in [l_{ij}, u_{ij}] \pm d(d + \sqrt{2}) \left[\sum_{m=1}^{b-2} (\sqrt{2} d)^{b-2-m} \epsilon_m \right] n^{b-1}. \tag{2.40}$$

By substituting the values of l_{ij} and u_{ij} from line 2(c), we have

$$t_{b-1}^{ij}(\rho_1, \ldots, \rho_n) \in e_{ij} \pm \left(d(d + \sqrt{2}) \left[\sum_{m=1}^{b-2} (\sqrt{2} d)^{b-2-m} \epsilon_m \right] + \epsilon_{b-1} \right) n^{b-1}. \tag{2.41}$$

[1] This holds even though on line 1 of Algorithm 2.12, we only estimate $d^2 |S|$ of the terms e_{ij} (i.e. EVAL does not actually estimate *all* terms in the recursive decomposition of t_k, as it does not need to) — this is because in our analysis of the probability bound for Algorithm 2.12, we actually produced a looser bound by assuming all n terms e_{ij} are estimated.

We conclude that

$$
\begin{aligned}
t_b(\rho_1,\ldots,\rho_n) &= \sum_{ij} \mathrm{Tr}(\sigma_j\rho_i) t^{ij}_{b-1}(\rho_1,\ldots,\rho_n) && (2.42)\\
&\subseteq \left[\sum_{ij} \mathrm{Tr}(\sigma_j\rho_i) e_{ij}\right] + && (2.43)\\
&\quad \left(d(d+\sqrt{2})\left[\sum_{m=1}^{b-2} (\sqrt{2}d)^{b-2-m}\epsilon_m\right] + \epsilon_{b-1}\right)\left[\sum_{ij}\mathrm{Tr}(\sigma_j\rho_i)\right] n^{b-1} && (2.44)\\
&\subseteq \left[\sum_{ij} \mathrm{Tr}(\sigma_j\rho_i) e_{ij}\right] + \\
&\quad \sqrt{2}d\left(d(d+\sqrt{2})\left[\sum_{m=1}^{b-2} (\sqrt{2}d)^{b-2-m}\epsilon_m\right] + \epsilon_{b-1}\right) n^b && (2.45)\\
&\subseteq \left[[L,U] \pm \epsilon_{b-1} d^2 n^b\right] + \sqrt{2}d\left(d(d+\sqrt{2})\left[\sum_{m=1}^{b-2} (\sqrt{2}d)^{b-2-m}\epsilon_m\right] + \epsilon_{b-1}\right) n^b \\
&\subseteq [L,U] \pm d(d+\sqrt{2})\left[\sum_{m=1}^{b-1} (\sqrt{2}d)^{b-1-m}\epsilon_m\right] n^b, && (2.46)
\end{aligned}
$$

where the third statement follows from a calculation similar to Equation (2.34), and the fourth statement from line 3(b) of Algorithm 2.14. This proves the claim of Equation (2.39).

To complete the proof of Lemma 2.16, observe that by Lemma 2.15, the assignment ρ^{opt} is feasible for P_2 with probability at least $1 - d^{2k}n^{k-f}$. Thus, plugging ρ^{opt} into each of the $d^2 n$ linear constraints produced by the recursive calls to LINEARIZE on each t^{ij}_{k-1}, we have by Equations (2.39) and (2.45) that (with probability $1 - d^{2k}n^{k-f}$) for $\mathrm{OPT}_P = t_k(\rho^{\mathrm{opt}})$,

$$
\begin{aligned}
t_k(\rho^{\mathrm{opt}}) &= \sum_{ij} \mathrm{Tr}\left(\sigma_j \rho_i^{\mathrm{opt}}\right) t^{ij}_{k-1}(\rho^{\mathrm{opt}}) && (2.47)\\
&\subseteq \left[\sum_{ij} \mathrm{Tr}(\sigma_j\rho_i^{\mathrm{opt}}) e_{ij}\right] \pm \sqrt{2}d\left(d(d+\sqrt{2})\left[\sum_{m=1}^{k-2} (\sqrt{2}d)^{k-2-m}\epsilon_m\right] + \epsilon_{k-1}\right) n^k \\
&\subseteq \mathrm{OPT}_2 \pm d(d+\sqrt{2})\left[\sum_{m=1}^{k-1} (\sqrt{2}d)^{k-1-m}\epsilon_m\right] n^k, && (2.48)
\end{aligned}
$$

where the last statement follows since ρ^{opt} is not necessarily the optimal solution to P_2. □

Acknowledgements for this chapter. We thank Jamie Sikora and Sarvagya Upadhyay for helpful feedback, and Yi-Kai Liu for interesting discussions. We wish to especially thank Oded Regev for many helpful comments and suggestions, and Richard Cleve for bringing our attention to the method of conditional expectations, and for stimulating discussions and support.

Chapter 3

Hardness of approximation for quantum problems

This chapter is based on [109]:

S. Gharibian and J. Kempe. Hardness of approximation for quantum problems. In *Proceedings of 39th International Colloquium on Automata, Languages and Programming*, pages 387-398, 2012, DOI: 10.1007/978-3-642-31594-7,

The polynomial hierarchy plays a central role in classical complexity theory. In this chapter, we define a quantum generalization of the polynomial hierarchy, and initiate its study. We show that not only are there natural complete problems for the second level of this quantum hierarchy, but that these problems are in fact hard to approximate. Using these techniques, we also obtain hardness of approximation for the class QCMA. Our approach is based on the use of dispersers, and is inspired by the classical results of Umans regarding hardness of approximation for the second level of the classical polynomial hierarchy [235]. We close the chapter by showing that two variants of the local Hamiltonian problem with hybrid classical-quantum ground states are complete and hard to approximate for the second level of our quantum hierarchy, respectively.

3.1 Introduction and results

Over the last decades, the Polynomial Hierarchy (PH) [193], a natural generalization of the class NP, has been the focus of much study in classical computational complexity. Of

particular interest is the second level of PH, denoted Σ_2^p. Here, we say a problem is in Σ_2^p if it has an efficient verifier with the property that for any YES instance $x \in \{0,1\}^n$ of the problem, *there exists* a polynomial length proof y such that *for all* polynomial length proofs z, the verifier accepts x, y and z. Note that the *alternation* from an existential quantifier over y to a for-all quantifier over z is crucial here – keeping only the existential quantifier reduces us to NP.

It turns out that introducing such alternating quantifiers makes Σ_2^p a powerful class believed to be *beyond* NP. For example, there exist natural and important problems known to be in Σ_2^p but not in NP. Such problems range from "does the optimal assignment to a 3SAT instance satisfy *exactly* k clauses?" to practically relevant problems related to circuit minimization, such as "given a boolean formula C in Disjunctive Normal Form (DNF), what is the smallest DNF formula C' equivalent to C?" (see, e.g. [235]). The study of Σ_2^p has also led to a host of other fundamental theoretical results, such as the Karp-Lipton theorem, which states that NP $\not\subseteq$ $\mathrm{P}_{/\mathrm{poly}}$ unless PH collapses to Σ_2^p. Σ_2^p has even been used to prove that SAT cannot be solved simultaneously in linear time and logarithmic space [98, 99]. For these reasons, Σ_2^p and more generally PH have occupied a central role in classical complexity theoretic research.

Moving to the quantum setting, the study of quantum proof systems and a natural quantum generalization of NP, the class Quantum Merlin Arthur (QMA) [171], has been a very active area of research over the last decade. Recall from Section 1.5.2 that a problem is in QMA if for any YES instance of the problem, there exists a polynomial size *quantum* proof convincing a quantum verifier of this fact with high probability. With the notion of quantum proofs in mind, we thus ask the natural question: *Can a quantum generalization of Σ_2^p be defined, and what types of problems might it contain and characterize?* Perhaps surprisingly, to date there are almost no known results in this direction.

Our results: In this chapter, we introduce a quantum generalization of Σ_2^p, which we call cq-Σ_2, and initiate its study. Our results include cq-Σ_2-completeness and cq-Σ_2-hardness of approximation for a number of new problems we define. Our techniques also yield hardness of approximation for the complexity class known as QCMA. We now describe these results in further detail.

Hardness of approximation for cq-Σ_2. To begin, we informally define cq-Σ_2 (see Section 3.2 for formal definitions).

Definition 3.1 (cq-Σ_2 (informal))**.** *A problem* Π *is in* cq-Σ_2 *if there exists an efficient quantum verifier satisfying the following property for any input* $x \in \{0,1\}^n$*:*

- *If* x *is a YES instance of* Π*, then there exists a* classical *proof* $y \in \{0,1\}^{\mathrm{poly}(n)}$ *such that for all* quantum *proofs* $|z\rangle \in \mathcal{B}^{\otimes \mathrm{poly}(n)}$*, the verifier accepts* x*,* y *and* $|z\rangle$ *with high probability.*
- *If* x *is a NO instance of* Π*, then for all* classical *proofs* $y \in \{0,1\}^{\mathrm{poly}(n)}$*, there exists a* quantum *proof* $|z\rangle \in \mathcal{B}^{\otimes \mathrm{poly}(n)}$ *such that the verifier rejects* x*,* y *and* $|z\rangle$ *with high probability.*

(Recall here that $\mathcal{B} := \mathbb{C}^2$.) We believe this is a natural quantum generalization of Σ_2^p. Here, the prefix *cq* in cq-Σ_2 follows since the existential proof is classical, while the for-all proof is quantum. One can also consider variations of this scheme such as qq-Σ_2, qc-Σ_2, or cc-Σ_2 (with a quantum verifier), defined analogously. In this chapter, however, our focus is on cq-Σ_2, as it is the natural setting for the computational problems for which we wish to prove hardness of approximation. Note also that unlike for Σ_2^p, the definition of cq-Σ_2 is bounded error – this is due to the use of a quantum verifier for cq-Σ_2. This implies, for instance, that the quantum analogue of the classically non-trivial result BPP $\subseteq \Sigma_2^p$ [227, 177], i.e. BQP $\subseteq$ cq-Σ_2, holds trivially. Finally, one can extend the definition of cq-Σ_2 to an entire hierarchy of quantum classes analogous to PH by adding further levels of alternating quantifiers, attaining presumably different classes depending on whether the quantifier at any particular level runs over classical or quantum proofs.

To next discuss hardness of approximation for cq-Σ_2, we recall two classical problems crucial to our work here. First, in the NP-complete problem SET COVER, one is given a set of subsets $\{S_i\}$ whose union covers a ground set U, and we are asked for the smallest number of the S_i whose union still covers U. If, however, the S_i are represented *succinctly* as the on-set[1] of a 3-DNF formula ϕ_i, we obtain a more difficult problem known as SUCCINCT SET COVER (SSC). SSC, along with a related problem IRREDUNDANT (IRR), are not just NP-hard, but are Σ_2^p-complete (indeed, they are even Σ_2^p-hard to approximate [235]). SSC and IRR are defined as:

Definition 3.2 (SUCCINCT SET COVER (SSC) [235])**.** *Given a set* $S = \{\phi_i\}$ *of* 3*-DNF formulae such that* $\bigvee_{i \in S} \phi_i$ *is a tautology, what is the size of the smallest* $S' \subseteq S$ *such that* $\bigvee_{i \in S'} \phi_i$ *a tautology?*

[1] By *on-set*, we mean the set of assignments which cause ϕ_i to be true.

Definition 3.3 (IRREDUNDANT (IRR) [235])**.** *Given a DNF formula $\phi = t_1 \vee t_2 \vee \cdots \vee t_n$, what is the size of the smallest $S \subseteq \{t_i\}_{i=1}^n$ such that $\phi \equiv \bigvee_{i \in S} t_i$?*

Our work introduces and studies quantum generalizations of SSC and IRR. In particular, analogous to the classically important task of circuit minimization, the quantum generalizations we define are arguably natural and related to what one might call "Hamiltonian minimization" – given a sum of Hermitian operators $H = \sum_i H_i$, what is the smallest subset of terms $\{H_i\}$ whose sum approximately preserves certain spectral properties of H? We hope that such questions may be useful to physicists in a lab who wish to simulate the simplest Hamiltonian possible while retaining the desired characteristics of a complex Hamiltonian involving many interactions. We remark that at a high level, the connection to cq-Σ_2 for the task of Hamiltonian minimization is as follows: The classical existential proof encodes the subset of terms $\{H_i\}$, while the quantum for-all proof encodes complex unit vectors which achieve certain energies against H. The problem QUANTUM SUCCINCT SET COVER is now defined as follows.

Definition 3.4. QUANTUM SUCCINCT SET COVER *(QSSC) (informal) Given a set of local Hamiltonians $\{H_i\}$ such that $\sum_i H_i$ has smallest eigenvalue at least α, what is the size of the smallest subset S of the H_i such that $\sum_{H_i \in S} H_i$ has smallest eigenvalue at least α? Any subset satisfying this property is called a* cover*.*

As defined in Section 1.5.4, a *local Hamiltonian* is a sum of Hermitian operators, each of which acts non-trivially on at most $k \in \Theta(1)$ qubits. Intuitively, the goal in QSSC is to cover the entire Hilbert space using as few interaction terms H_i as possible. Hence, we associate the notion of a "cover" with obtaining large eigenvalues, as opposed to small ones, making QSSC a direct quantum analogue of SSC. We remark that since SSC is a classical constraint satisfaction problem, we believe the language of *quantum* constraint satisfaction, i.e. Hamiltonian constraints, is a natural avenue for defining QSSC. Our first result concerns QSSC, and is as follows.

Theorem 3.5. *QSSC is* cq-Σ_2*-complete, and moreover is* cq-Σ_2*-hard to approximate within $N^{1-\epsilon}$ for all $\epsilon > 0$, where N is the encoding size of the QSSC instance.*

By *hard to approximate*, we mean that any problem in cq-Σ_2 can be reduced to an instance of QSSC via a polynomial time mapping or Karp reduction such that the gap between the sizes of the optimal cover in the YES and NO cases scales as $N^{1-\epsilon}$. In other words, it is cq-Σ_2-hard to determine whether the smallest cover size of an arbitrary instance of QSSC is at most g or at least g' for $g'/g \in \Omega(N^{1-\epsilon})$ (where $g' \geq g$). We next define the problem QUANTUM IRREDUNDANT (QIRR).

Definition 3.6. QUANTUM IRREDUNDANT *(QIRR) (informal) Given a set of succinctly described orthogonal projection operators* $\{H_i\}$ *acting on* N *qubits, and* $\{c_i \geq 0\} \subseteq \mathbb{R}$*, define* $H := \sum_i c_i H_i$*. Then, what is the size of the smallest subset* $S \subseteq \{H_i\}$ *such that for* $H' = \sum_{H_i \in S} c_i H_i$*, vectors achieving high and low energies against* H *continue to obtain high and low energies against* H'*, respectively?*

Here, by a *succinctly* described projector, we mean a possibly non-local operator which is the tensor product of k-local projectors for some $k \in \Theta(1)$. This non-local structure naturally generalizes IRR, where the DNF formula is allowed to be non-local. Our next result is the following.

Theorem 3.7. *QIRR is* cq-Σ_2*-hard to approximate within* $N^{\frac{1}{2}-\epsilon}$ *for all* $\epsilon > 0$*, where* N *is the encoding size of the QIRR instance.*

Hardness of approximation for QCMA. The techniques from above can also be used to show hardness of approximation for QCMA. Here, the class QCMA [22] is defined as cq-Σ_2 with the second (quantum) proof omitted, and can hence be thought of as the first level of our "*cq*-hierarchy". By defining the problem QUANTUM MONOTONE MINIMUM SATISFYING ASSIGNMENT (QMSA) (see Section 3.5), we show:

Theorem 3.8. *QMSA is QCMA-complete, and moreover is QCMA-hard to approximate within* $N^{1-\epsilon}$ *for all* $\epsilon > 0$*, where* N *is the encoding size of the QMSA instance.*

A canonical cq-Σ_2**-complete problem.** Our last results the canonical Σ_2^p-complete problem Σ_iSAT and its generalization to the quantum setting. Specifically, given a boolean formula ϕ, Σ_iSAT asks whether:

$$\exists \mathbf{x}_1 \forall \mathbf{x}_2 \exists \mathbf{x}_3 \cdots \forall \mathbf{x}_i \;\text{ such that }\; \phi(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_i) = 1. \tag{3.1}$$

Here, we have assumed i is even; for odd i, the last quantifier is a $\exists$. The terms $\mathbf{x}_j$ are vectors of boolean variables. For $i = 2$, one can define a natural quantum generalization of this problem, denoted cq-Σ_2LH and defined in Section 3.6, using local Hamiltonians whose ground states are tensor products of a classical string and a quantum state. We show:

Theorem 3.9. cq-Σ_2LH *is* cq-Σ_2*-complete.*

Moreover, by defining an appropriate variant of cq-Σ_2LH, denoted cq-Σ_2LH-HW and also defined in Section 3.6, where the goal is to minimize the Hamming weight of the classical portion of the ground states mentioned above, we obtain the following result.

Theorem 3.10. cq-Σ_2LH-HW *is* cq-Σ_2*-complete, and moreover is* cq-Σ_2*-hard to approximate within* $N^{1-\epsilon}$ *for any* $\epsilon > 0$*, for* N *the encoding size of the* cq-Σ_2LH-HW *instance.*

Proof ideas: Our proofs are inspired by the classical work of Umans [235, 219], and are achieved in a few steps. First, we show a *gap-introducing* reduction from an arbitrary cq-Σ_2 problem to a problem we call QUANTUM MONOTONE MINIMUM WEIGHT WORD (QMW) using *dispersers* (see e.g., [228, 232]). We then show the following *gap-preserving* reductions, where $\leq_K$ denotes a mapping or Karp reduction:

$$\text{QMW} \leq_K \text{QSSC} \leq_K \text{QIRR} \ . \tag{3.2}$$

This yields hardness ratios of N^ϵ for some $\epsilon > 0$. To obtain the stronger results claimed in Section 3.1, we finally apply the gap amplification of Umans [235] and improved disperser construction of Ta-Shma, Umans, and Zuckerman [232].

In the classical setting, Umans [235, 219] used dispersers to attain hardness of approximation results relative to Σ_2^p for the classical problems MMWW (the classical version of QMW), SSC and IRR. To extend his techniques to the quantum setting, the most involved aspects of our work are the gap-preserving reductions from QMW to QSSC to QIRR. Here, an intricate balancing act involving carefully defined local Hamiltonian terms is needed to construct operators with the spectral properties required for our reductions. To analyze the resulting sums of non-commuting Hamiltonians, we require heavier machinery, such as the specific structure of Kitaev's local Hamiltonian construction [171], the Projection Lemma of Kempe, Kitaev, and Regev [163], and the Geometric Lemma of Kitaev [171].

Finally, to show cq-Σ_2-completeness of cq-Σ_2LH, we study the interplay between proofs of a classical-quantum structure and Kempe and Regev's [164] 3-local Hamiltonian construction. Specifically, a careful analysis reveals that any cq-Σ_2 verification circuit can be modified in such a way that fixing the value c of its classical proof register leads to an *effective* Hamiltonian H_c. We then study the spectrum of H_c to achieve the desired result. Moving on to cq-Σ_2LH-HW, hardness of approximation is now attained by combining our reduction for cq-Σ_2LH with the result that QMW is hard to approximate.

Previous and related work: In terms of hardness of approximation, the related question of whether a *quantum* PCP theorem holds is currently one of the biggest open problems in quantum complexity theory (see, e.g., [6, 17, 26, 134]). Regarding quantum generalizations of PH, the only previous work we are aware of is that of Yamakami [260]. However, the results of Yamakami are largely unrelated to ours (for example, complete problems are

not studied), and the proposed definition of Reference [260] differs from ours in a number of ways: It is based on quantum Turing machines (whereas we work with quantum circuits), allows *quantum* inputs (whereas here, like QMA, the input to a problem is a classical string), and considers quantum quantifiers at each level of the hierarchy (whereas in its full generality our scheme allows alternating between classical and quantum quantifiers between levels as desired).

Significance and open questions: The classical polynomial hierarchy plays an important role in classical complexity theory, both as a generalization of NP and as a proof tool in itself. It is hoped that the scheme we propose here for generalizing PH to the quantum setting will find similar applications in quantum complexity theory. Second, the problems we show to be cq-Σ_2-complete here are arguably natural, and in embodying a generalization of classical circuit minimization or optimization, may hopefully be related to practical scenarios in a lab. Further, although the alternation between classical and quantum quantifiers in cq-Σ_2 may *a priori* seem odd, the notion of relating a classical proof to, say, subsets of local Hamiltonian terms, and the quantum proof to quantum states achieving certain energies is in itself quite natural, and in our opinion justifies the study of such a combination of quantifiers. Third, with respect to hardness of approximation, since whether a quantum PCP theorem holds remains a challenging open question, it is all the more interesting that one is able to prove hardness of approximation in a quantum setting here using an entirely different tool, namely that of dispersers. We remark that dispersers and their two-sided analogues, extractors, have been used classically to amplify existing PCP inapproximability results [228, 263]. However, as far as we are aware, neither are known to directly yield PCP constructions.

We leave a number of questions open: What other natural problems are complete for cq-Σ_2 or higher levels? Can we say anything non-trivial about the relationship between Σ_2^p and cq-Σ_2? How do the different classes cq-Σ_2, qc-Σ_2, qq-Σ_2, and cc-Σ_2 relate to each other? Where do the quantum hierarchies obtained by extending cq-Σ_2 to higher levels sit relative to known complexity classes? We hope the answers to such questions will help establish classes like cq-Σ_2 as fundamental concepts in the setting of quantum computational complexity.

Organization of this chapter: We begin in Section 3.2 by formally defining the classes and problems studied in this chapter. In Section 3.3, we prove that QSSC and QIRR are hard to approximate for cq-Σ_2 within N^ϵ; this is further improved in Section 3.4. Section 3.5 presents hardness of approximation results for QCMA. We close in Section 3.6 by showing

cq-Σ_2-completeness of cq-Σ_2LH and cq-Σ_2-hardness of approximation for cq-Σ_2LH-HW.

3.2 Definitions

We now define relevant classes and problems, and state lemmas which prove useful in our analysis. Throughout our discussion, recall that $\mathcal{B} := \mathbb{C}^2$, and for a set S of matrices over $\mathbb{C}$, let $H_S := \sum_{H_i \in S} H_i$.

We begin with a formal definition of cq-Σ_2. Recall that a promise problem is a pair $A = (A_{\text{yes}}, A_{\text{no}})$ such that $A_{\text{yes}}, A_{\text{no}} \subseteq \{0,1\}^*$ and $A_{\text{yes}} \cap A_{\text{no}} = \emptyset$.

Definition 3.11 (cq-Σ_2)**.** *Let $A = (A_{\text{yes}}, A_{\text{no}})$ be a promise problem. We say that $A \in$ cq-Σ_2 if there exist polynomially bounded functions $t, c, q : \mathbb{N} \mapsto \mathbb{N}$, and a deterministic Turing machine M acting as follows. For every n-bit input x, M outputs in time $t(n)$ a description of a quantum circuit V_x such that V_x takes in a $c(n)$-bit proof $|c\rangle$, a $q(n)$-qubit proof $|q\rangle$, and outputs a single qubit. We say V_x* accepts *$|c\rangle|q\rangle$ if measuring its output qubit in the computational basis yields* 1*. Then:*

- *Completeness: If $x \in A_{\text{yes}}$, then $\exists\ |c\rangle$ such that $\forall\ |q\rangle$, V_x accepts $|c\rangle|q\rangle$ with probability $\geq 2/3$.*
- *Soundness: If $x \in A_{\text{no}}$, then $\forall\ |c\rangle$, $\exists\ |q\rangle$ such that V_x rejects $|c\rangle|q\rangle$ with probability $\geq 2/3$.*

Note that the completeness and soundness parameters can be amplified to values exponentially close to 1. Specifically, we use the standard approach of repeating V_x polynomially many times in parallel (see "Error reduction for QMA" in Section 1.5.2), except that we only need one copy of the classical register $\mathcal{C}$ for all parallel runs. For any value c placed in $\mathcal{C}$, we think of it as being "hardwired" into V_x, thus obtaining a quantum verification circuit $V_{x,c}$, which we now apply in parallel to the many copies of the quantum proof. The standard weak error reduction analysis for QMA now applies (see, e.g. [22]). Throughout this chapter, we refer to this as *error reduction*.

We next define the terms cQMA circuit, monotone set, QMW, QSSC, and QIRR.

Definition 3.12 (cQMA circuit)**.** *Let $n, m \in \mathbb{N}^+$. A* cQMA *circuit V is a quantum circuit receiving n bits in an* INPUT *register and m qubits in a* CHOICE *register, and outputting a single qubit $|a\rangle$. We say:*

- V accepts $x \in \{0,1\}^n$ *in INPUT if for all* $|y\rangle \in \mathcal{B}^{\otimes m}$ *in CHOICE, measuring* $|a\rangle$ *in the computational basis yields* 1 *with probability at least* $2/3$.
- V rejects $x \in \{0,1\}^n$ *in INPUT if there exists a* $|y\rangle \in \mathcal{B}^{\otimes m}$ *in CHOICE such that measuring* $|a\rangle$ *in the computational basis yields* 0 *with probability at least* $2/3$.

Definition 3.13 (Monotone set)**.** *A set* $S \subseteq \{0,1\}^n$ *is called* monotone *if for any* $x \in S$*, any string obtained from* x *by flipping one or more zeroes in* x *to one is also in* S.

Definition 3.14 (QUANTUM MONOTONE MINIMUM WEIGHT WORD (QMW))**.** *Given a* cQMA *circuit* V *accepting exactly a non-empty monotone set* $S \subseteq \{0,1\}^n$*, and integer thresholds* $0 \leq g \leq g' \leq n$*, output:*

- *YES if there exists an* $x \in \{0,1\}^n$ *of Hamming weight at most* g *accepted by* V.
- *NO if all* $x \in \{0,1\}^n$ *of Hamming weight at most* g' *are rejected by* V.

Note that clearly QMW $\in$ cq-Σ_2.

Definition 3.15 (QUANTUM SUCCINCT SET COVER (QSSC))**.** *Let* $S := \{H_i\}$ *be a set of* 5*-local Hamiltonians* H_i *acting on* N *qubits such that* $\sum_{H_i \in S} H_i \succeq \alpha I$ *for* $\alpha > 0$*. Then, given* $\beta \in \mathbb{R}$ *such that* $\alpha - \beta \geq 1$ *and integer thresholds* $0 \leq g \leq g'$*, output:*

- *YES if there exists* $S' \subseteq S$ *of cardinality at most* g *such that* $\sum_{H_i \in S'} H_i \succeq \alpha I$.
- *NO if for all* $S' \subseteq S$ *of size at most* g'*,* $\sum_{H_i \in S'} H_i$ *has an eigenvalue at most* β.

Any S' *satisfying the YES case is called a* cover.

Note that requiring $\alpha - \beta \in \Omega(1)$ above is without loss of generality, as any instance of QSSC with gap $1/p(N)$ for p a polynomially bounded function can be modified to obtain an equivalent instance with constant gap by multiplying each H_i by $p(N)$ [248] (see Section 1.5.4).

Definition 3.16 (QUANTUM IRREDUNDANT (QIRR))**.** *Given* $S := \{c_i H_i\}$*, where each* H_i *acts on* N *qubits and is a tensor product of* 5*-local orthogonal projection operators and* $c_i \geq 0$ *are real. Then, given* $\alpha, \beta \in \mathbb{R}$ *such that* $\alpha - \beta \geq 1$*, and integer thresholds* $0 \leq g \leq g'$*, output:*

- *YES if there exists* $S' \subseteq S$ *of cardinality at most* g *such that for all* $|\psi\rangle \in \mathcal{B}^{\otimes N}$*:*

  - *If* $\mathrm{Tr}(H_S|\psi\rangle\langle\psi|) \geq \alpha$, *then* $\mathrm{Tr}(H_{S'}|\psi\rangle\langle\psi|) \geq \alpha$, *and*
  - *If* $\mathrm{Tr}(H_S|\psi\rangle\langle\psi|) \leq \beta$, *then* $\mathrm{Tr}(H_{S'}|\psi\rangle\langle\psi|) \leq \beta$.
- *NO if for all* $S' \subseteq S$ *of cardinality at most* g', *there exists a state* $|\psi\rangle \in \mathcal{B}^{\otimes N}$ *with* $\mathrm{Tr}(H_S|\psi\rangle\langle\psi|) \geq \alpha$ *and* $\mathrm{Tr}(H_{S'}|\psi\rangle\langle\psi|) \leq \beta$.

Roughly, QSSC asks how many local interaction terms in a local Hamiltonian one can discard while maintaining the value of the worst assignment. This is intended to mimic the idea of maintaining a tautology for a 3-DNF formula in SSC classically. Analogous to the relationship between SSC and IRR, QIRR allows possibly non-local Hamiltonian terms so long as they have a succinct description (this generalizes the use of superconstant arity in IRR) and are projectors up to scalar multiplication (this generalizes the requirement that each term t_i in IRR is an AND of variables). QIRR then asks how many interaction terms can be discarded in a sum of such Hamiltonian terms while ensuring that any assignment $|\psi\rangle$ achieves approximately the same value on both the original and modified Hamiltonians.

Next, the key tool enabling the creation of a gap in our reductions is a *disperser* (see e.g. [228, 232]).

Definition 3.17 (Disperser)**.** *Let* $G = (L, R, E)$ *be a bipartite graph with* $|L| = 2^n$, $|R| = 2^m$ *and left-degree* 2^d. *Then,* G *is called a* (k, ϵ)-disperser *if, for any subset* $L' \subseteq L$ *of size* $|L'| \geq 2^k$, L' *has at least* $(1-\epsilon)|R|$ *neighbors in* R. *Moreover, if for any pair* (v, i) *for* $v \in L$, *one can compute the* i*th neighbor of* v *in time polynomial in* n, *then the disperser is called* explicit.

Finally, in this chapter we use the following useful known facts from local Hamiltonian complexity theory. To begin, we have two lemmas used to bound the eigenvalues of a pair of non-commuting operators. The first of these is the Geometric Lemma of Kitaev, which we stated as Lemma 1.8 in Section 1.5.5. The second is the Projection Lemma, stated below.

Lemma 3.18 (Kempe, Kitaev, Regev [163], Projection Lemma)**.** *Let* $Y = Y_1 + Y_2$ *act on Hilbert space* $\mathcal{H} = \mathcal{S} + \mathcal{S}^\perp$ *for Hamiltonians* Y_1 *and* Y_2. *Denote the zero eigenspace of* Y_2 *as* $\mathcal{S}$, *and assume the* Y_2 *eigenvectors in* $\mathcal{S}^\perp$ *have eigenvalue at least* $J > 2\,\|\,Y_1\,\|_\infty$. *Then, for* $\lambda(Y)$ *the smallest eigenvalue of* Y *and* $Y|_{\mathcal{S}} := \Pi_{\mathcal{S}} Y \Pi_{\mathcal{S}}$,

$$\lambda(Y_1|_{\mathcal{S}}) - \frac{\|\,Y_1\,\|_\infty^2}{J - 2\,\|\,Y_1\,\|_\infty} \leq \lambda(Y) \leq \lambda(Y_1|_{\mathcal{S}}) \ . \tag{3.3}$$

We next briefly review the elements of Kitaev's circuit-to-Hamiltonian construction [171] which play an important role in this chapter (see in Section 1.5.5 for an in-depth treatment). Given a cq-Σ_2 verification circuit $V = V_L \cdots V_1$ (where without loss of generality, each V_i is a one- or two-qubit unitary) acting on n proof bits (register A), m proof qubits (register B), and p ancilla qubits (register C), recall that this construction outputs a 5-local Hamiltonian H acting on $A \otimes B \otimes C \otimes D$, where D is a clock register consisting of L qubits. We then have $H := H_{\text{in}} + H_{\text{out}} + H_{\text{prop}} + H_{\text{stab}}$, for *penalty* terms as defined below:

$$H_{\text{in}} := I_{A,B} \otimes \left(\sum_{i=1}^{p} |1\rangle\langle 1|_{C_i} \right) \otimes |0\rangle\langle 0|_D \tag{3.4}$$

$$H_{\text{out}} := I_A \otimes |0\rangle\langle 0|_{B_1} \otimes I_C \otimes |L\rangle\langle L|_D \tag{3.5}$$

$$H_{\text{prop}} := \sum_{j=1}^{L} H_j, \text{ where } H_j \text{ is defined as} \tag{3.6}$$

$$-\frac{1}{2} V_j \otimes |j\rangle\langle j-1|_D - \frac{1}{2} V_j^\dagger \otimes |j-1\rangle\langle j|_D + \frac{1}{2} I \otimes (|j\rangle\langle j| + |j-1\rangle\langle j-1|)_D \tag{3.7}$$

$$H_{\text{stab}} := I_{A,B,C} \otimes \sum_{i=1}^{L-1} |01\rangle\langle 01|_{D_i, D_{i+1}}. \tag{3.8}$$

Above, the notation A_i refers to the ith qubit of register A (similarly for B, C, D). For any prospective proof $|\psi\rangle$ in $\text{Tr}(H|\psi\rangle\langle\psi|)$, each penalty term has the following effect on the structure of $|\psi\rangle$: H_{in} ensures that at time zero, the ancilla register is set to zero as it should be for V. H_{out} ensures that at time step L of V, measuring the output qubit causes acceptance with high probability. H_{prop} forces all steps of V appear in superposition in $|\psi\rangle$ with equal weights. Finally, note that for H_{in}, H_{out}, and H_{prop} above, time t in clock register D is implicitly encoded in unary as $\left|1^t 0^{L-t}\right\rangle$ (for H_{stab} above, register D is already explicitly written in unary); H_{stab} is thus needed to prevent invalid encodings of time steps from appearing in D.

We use two important properties of this construction. First, the null space of $H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}}$ is the space of *history states*, which for arbitrary $|\psi\rangle_{A,B}$ are defined as

$$|\psi\rangle_{\text{hist}} := \frac{1}{\sqrt{L+1}} \sum_{i=0}^{L} V_i \cdots V_1 |\psi\rangle_{A,B} \otimes |0\rangle_C \otimes |i\rangle_D. \tag{3.9}$$

For cq-Σ_2 circuits V, it is convenient to define for $c \in \{0,1\}^n$ and $|q\rangle \in \mathcal{B}^{\otimes m}$ the shorthand $|c, q\rangle_{\text{hist}} := |\psi\rangle_{\text{hist}}$ for $|\psi\rangle = |c\rangle|q\rangle$. The second important property of H we use is that its spectrum is related to V as follows.

Lemma 3.19 (Kitaev [171]). *The construction above maps V to (H, a, b) satisfying:*

- *If there exists a proof $|\psi\rangle$ accepted by V with probability at least $1-\epsilon$, then $|\psi\rangle_{\text{hist}}$ achieves $Tr(H|\psi\rangle\langle\psi|_{\text{hist}}) \leq a$ for $a := \epsilon/(L+1)$.*
- *If V rejects all proofs $|\psi\rangle$, then $H \succeq bI$ for $b \in \Omega\left(\frac{1-\sqrt{\epsilon}}{L^3}\right)$.*

3.3 Hardness of approximation for cq-Σ_2

We now show hardness of approximation for cq-Σ_2 for the problems QMW, QSSC, and QIRR. We begin with a gap-introducing reduction from an arbitary problem in cq-Σ_2 to QMW. We remind the reader that the hardness ratios obtained here are further strengthened in Section 3.4.

Theorem 3.20. *There exists a polynomial time reduction which, given an instance of an arbitrary* cq-Σ_2 *problem, outputs an instance of QMW with thresholds g and g' satisfying $g'/g \in \Theta(N^\epsilon)$ for some $\epsilon > 0$, where N is the encoding size of the QMW instance.*

Proof. The reduction follows Theorem 1 of Umans [235] closely; the points where we deviate from [235] are explicitly noted. Let Π be an instance of an arbitrary promise problem $A = (A_{\text{yes}}, A_{\text{no}})$ in cq-Σ_2 with encoding size n, and whose verification circuit V has a $c(n)$-bit existential proof register and a $q(n)$-qubit for-all proof register. We wish to map Π to a cQMA circuit W for QMW such that W accepts strings of small or large Hamming weight depending on whether $\Pi \in A_{\text{yes}}$ or $\Pi \in A_{\text{no}}$, respectively. To do so, we follow [235] and construct an explicit $(k, 1/2)$-disperser $G = (L, R, E)$ with left-degree 2^d using Reference [228], where $|L| = 2^{c(n)+1}$, $|R| = 2^{k+d-O(1)}$, and $k := \gamma \log c(n)$ for $\gamma \in \Theta(1)$ to be set as needed. Note that the value of d depends on the specific disperser construction used — for the construction of [228], we have $d = 4k + O(\log n)$. Roughly, the idea of Umans is now to have L correspond to assignments for the $c(n)$-bit classical register of V, and R to assignments for the classical register of W (in the setting of [235], note that W is a classical circuit). We then *encode* assignments from L by instead choosing neighbor sets in R. By exploiting the properties of dispersers, one can ensure that the sizes of the neighbor sets in R chosen vary widely between YES and NO cases for Π.

Specifically, imagine the vertices in L are arranged into a complete binary tree whose $2^{c(n)}$ leaves denote the $2^{c(n)}$ possible assignments to V's classical register. For convenience, we henceforth use L to mean this tree. Now, let $x \in \{0,1\}^{c(n)}$ denote a leaf of L. Then, a

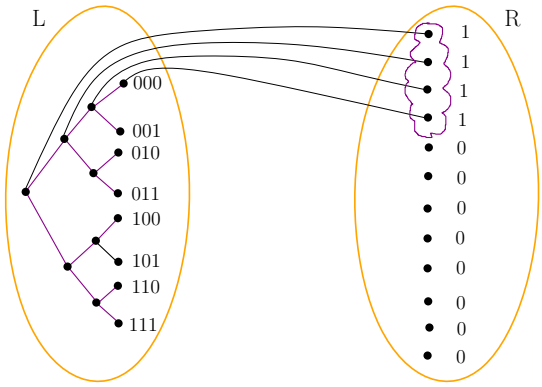


Figure 3.1: Here, the string $11110\cdots0$ in R encodes the string 000 in L. (Note: This graph is not a disperser, but nevertheless illustrates the encoding scheme.)

subset of vertices $R' \subseteq R$ is said to *encode* x if it contains the union of the neighbor sets of all vertices in the unique path from the root of L to x. Figure 3.1 illustrates this encoding scheme. How do the vertices of R then relate to W? Each vertex $r \in R$ corresponds to an input bit of W – setting this rth bit to one means we "choose" vertex r.

With the encoding scheme defined, we now construct the cQMA circuit W. Given y and $|z\rangle$ to its INPUT and CHOICE registers, respectively, W acts as follows: (a) If y corresponds to a subset $R_y \subseteq R$ such that $|R_y| > |R|/2$, then W sets its output qubit to one. (b) If $|R_y| \leq |R|/2$, then W first *decodes* R_y to obtain the set of leaves $L_y \subseteq L$. Roughly, it then outputs one if there exists $x \in L_y$ causing Π's verification circuit V to output one when fed the proofs x and $|z\rangle$. These last two steps require further clarification, which we now provide.

First, given $R_y \subseteq R$, decoding it to obtain the set of leaves $L_y \subseteq L$ might *a priori* require exponential time, as recall $|L| = 2^{c(n)+1}$. This, however, is precisely where dispersers play their part: Since we set $\epsilon = 1/2$ in constructing our disperser, we know that for any $S \subseteq R$ with $|S| \leq |R|/2$, there are at most $2^k = c(n)^\gamma$ vertices in L whose neighbor sets are completely contained in S. Thus, by starting at the root of L and performing a breadth-first-search down the tree (where we prune any branches along which we encounter a vertex whose neighbor set is not contained in R_y, as by definition such vertices cannot encode any leaf x), we can efficiently decode R_y to obtain L_y while visiting only polynomially vertices in L. It remains to specify how W checks whether there exists an $x \in L_y$ causing V to accept, and here we must deviate from Umans' construction.

First, if $|L_y| = 1$, our task is straightforward – simply run V as a black box on proofs $x \in L_y$ and $|z\rangle$, and output the result. Then, W outputs one with probability at least

2/3 on input y for all quantum proofs $|z\rangle$ if and only if V also does so on proofs x and $|z\rangle$. If , however, $|L_y| > 1$, a more involved construction of W is necessary. Here, W takes three inputs: a classical description of V, an $|R|$-bit string y to denote subsets in R, and a $2^k q(n)$-qubit proof $|z\rangle$. Then, for the ith candidate string $x_i \in L_y$, W feeds x_i and the ith block of $q(n)$ proof qubits of $|z\rangle$ into V. (If $|L_y| < 2^k$, we simply re-use values of $x \in L_y$ in the leftover parallel runs of V.) W then coherently computes the OR of the output qubits of all parallel runs of V and outputs this qubit as its answer.

Let us briefly justify why this works. For simplicity, assume the quantum proof to W can be written $|z\rangle = |z_1\rangle \otimes \cdots \otimes |z_{2^k}\rangle$; entangled proofs can be shown not to pose a problem via the same proof technique used in standard error reduction [22]. Now, if there exists an $x_i \in L_y$ causing V to accept for all quantum proofs, then in the ith parallel run of V in W corresponding to x_i, V outputs 1 with probability at least 2/3 on any $|z_i\rangle$, implying W outputs 1 with probability at least 2/3. Conversely, if for all $x_i \in L_y$, there exists a quantum proof $|z_i\rangle$ rejected by V, then by standard error reduction for V and the union bound, the state $|z\rangle = |z_1\rangle \otimes \cdots \otimes |z_{2^k}\rangle$ causes W to output 1 with probability at most 1/3, as required.

Following Reference [235] again, we now argue that W accepts a *non-empty monotone* set, and we analyze the hardness gap introduced by this reduction. The first of these is simple – namely, W accepts a set $R' \subseteq R$ if either $|R| > |R/2|$, in which case it also accepts any $R'' \supseteq R'$, or if R' encodes some $x \in L$ accepted by V, in which case any $R'' \supseteq R'$ would also encode x and hence be accepted. As for the gap, if $x \in L$ is an accepting assignment for V when $\Pi \in A_{\text{yes}}$, then to encode x using a subset of R requires at most $c(n)2^d$ vertices in R, where recall 2^d is the left-degree of our disperser. On the other hand, if $\Pi \in A_{\text{no}}$, then the only way for W to accept is to choose $R' \subseteq R$ with $|R'| > |R|/2 \approx c(n)^\gamma 2^d$. This yields a hardness ratio of $\Omega(c(n)^{\gamma-1})$. Since W's encoding size N is polynomial in $c(n)$, there exists some $\epsilon > 0$ such that the ratio produced is of order N^ϵ, as desired. □

We next show a gap-preserving reduction from QMW to QSSC. Its proof requires Lemmas 3.22 and 3.23, which are stated and proven subsequently.

Theorem 3.21. *QSSC is in* cq-Σ_2. *Further, there exists a polynomial time reduction which, given an instance of QMW with thresholds f and f', outputs an instance of QSSC with thresholds $g = f + 2$ and $g' = f' + 2$, respectively.*

Proof. That QSSC is in cq-Σ_2 follows using Kitaev's verifier [171] for putting k-local Hamiltonian in QMA. Specifically, we construct a cq-Σ_2 verification circuit for QSSC which takes a description c of some subset of local Hamiltonians $S := \{H_i\}$ in its classical register, and

estimates the energy achieved by $|q\rangle$ in its quantum register against H_S using Kitaev's approach, outputting zero or one according to whether the measured energy is above or below the desired thresholds.

To reduce QMW to QSSC, suppose we are given a cQMA circuit V accepting exactly a non-empty monotone set $T \subseteq \{0,1\}^n$ and threshold parameters f and f'. We assume without loss of generality that V is represented as a sequence of one and two qubit unitary gates V_i such that $V = V_L \cdots V_1$. We also assume using standard error reduction that if V accepts (rejects) input $x \in \{0,1\}^n$, then it outputs one (zero) with probability at least $1 - \epsilon := 1 - 2^{-4(n+m)}$.

We now state our instance $(S, \alpha, \beta, g, g')$ of QSSC as follows. We first apply Kitaev's circuit-to-Hamiltonian construction from Section 3.2 to V to obtain a 3-tuple (H, a, b). Note that $H = \sum_{i=1}^{r} H_i$ with r terms $0 \preceq H_i \preceq I$. Then, set $\alpha := 1 - (\zeta + 1)\epsilon$, and $\zeta := 2(1 + 2^{2(n+m)})/(L+1)$. Define $\beta := 1 - b$. Note that for large $n + m$, this yields $\alpha \geq 1 - 2^{-(n+m)}$ and $\beta \leq 1 - c(1 - 2^{-(n+m)})/L^3$ for some constant c. Further, define $g := f + 2$, $g' := f' + 2$, and let S consist of the elements (intuition to follow)

$$\begin{aligned} G_1 &:= (L+1)|0\rangle\langle 0|_{A_1} \otimes I_{B,C} \otimes |0\rangle\langle 0|_D && (3.10) \\ &\vdots \\ G_n &:= (L+1)|0\rangle\langle 0|_{A_n} \otimes I_{B,C} \otimes |0\rangle\langle 0|_D && (3.11) \\ G_{n+1} &:= (\Delta+1)(H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}}) && (3.12) \\ G_{n+2} &:= I - (H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}} + H_{\text{out}}), && (3.13) \end{aligned}$$

for $\Delta \geq 0$ to be chosen as required, and where A_i denotes the ith qubit of register A. Intuitively, the terms in S play the following roles: G_{n+1} penalizes assignments which are not valid history states. G_{n+2} penalizes valid history states accepted by V. Finally, the G_i for $i \in [n]$ penalize valid history states rejected by V (recall that V accepts a monotone set, and so flipping a one to a zero in register A may lead V to reject). Thus, we cover the entire space. We now make this rigorous.

As required by Definition 3.15, we begin by showing that S itself is a cover, i.e. that $G_S \succeq \alpha I_{A,B,C,D}$. First, note that

$$G_S = I + \sum_{i=1}^{n} G_i - H_{\text{out}} + \Delta(H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}}). \tag{3.14}$$

It thus suffices to prove that for large enough Δ,

$$\Delta(H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}}) + \left(\sum_{i=1}^{n} G_i\right) - H_{\text{out}} \succeq -(\zeta+1)\epsilon I. \tag{3.15}$$

To show this, we use Lemma 3.18, the Projection Lemma, with

$$Y_1 := \left(\sum_{i=1}^{n} G_i\right) - H_{\text{out}}, \qquad Y_2 := \Delta(H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}}). \tag{3.16}$$

Intuitively, the Projection Lemma tells us that by increasing our weight Δ, we can force the smallest eigenvalue of $Y_1 + Y_2$ to be approximately the smallest eigenvalue of Y_1 restricted to the null space of Y_2. In our setting, this implies it suffices to study the smallest eigenvalue of Y_1 restricted to the space of all *valid* history states, i.e. states of the form of Equation (3.9). Let $\mathcal{S}_{\text{hist}}$ denote the space of valid history states; note $\mathcal{S}_{\text{hist}}$ is the null space of $H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}}$. Then, in the notation of Lemma 3.18, to lower bound $\lambda(Y_1|_{\mathcal{S}_{\text{hist}}})$, we invoke Lemma 3.23 to instead upper bound the largest eigenvalue of $(-Y_1)|_{\mathcal{S}_{\text{hist}}}$. This yields $\lambda(Y_1|_{\mathcal{S}_{\text{hist}}}) \geq -\zeta\epsilon$. Noting that $\| Y_1 \|_\infty \leq n(L+1)+1$, and since by Lemma 3.22 the smallest non-zero eigenvalue of Y_2 scales as $\Omega(\Delta/L^3)$, it follows by Lemma 3.18 that by setting $\Delta \in \Omega(n^2 L^5/\epsilon)$, we have $Y_1 + Y_2 \succeq -(\zeta+1)\epsilon I$, as desired. This completes the proof that S is a cover.

We now show the desired reduction. Assume first that V accepts a string x of Hamming weight k, and let $T \subseteq [n]$ be such that $i \in T$ if and only if $x_i = 1$. We claim there exists a cover $S' \subseteq S$ of size $|S'| = k+2$ which consists of G_{n+1}, G_{n+2}, and the k terms G_i such that $i \in T$. To show this, following the proof above, the analogue of Equation (3.15) which we must prove is

$$\Delta(H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}}) + \left(\sum_{i \in T} G_i\right) - H_{\text{out}} \succeq -(\zeta+1)\epsilon I. \tag{3.17}$$

First, applying Lemma 3.23 again, we lower bound the smallest eigenvalue of

$$Y_1' := \left(\sum_{i \in T} G_i\right) - H_{\text{out}} \tag{3.18}$$

restricted to $\mathcal{S}_{\text{hist}}$ by $-\zeta\epsilon$. Since $\| Y_1' \|_\infty \leq \| Y_1 \|_\infty$ for Y_1 from the previous case of $T = [n]$, the value of Δ from before still suffices to apply Lemma 3.18 and conclude that Equation (3.17) holds, as desired.

Conversely, suppose V rejects any string x of Hamming weight at most k. For any $S' \subseteq S$ with $|S'| \leq k+2$, we claim that $G_{S'}$ has an eigenvalue at most β. To see this, note first that if $G_{n+2} \notin S'$, then the state $|1^n, y\rangle_{\text{hist}}$ attains expected value zero against $G_{S'}$, where note $\beta \geq 0$. Similarly, if $G_{n+1} \notin S'$, then the state $|1^n\rangle_{A,B,C} \otimes |0\rangle_D$ obtains expected

value at most zero against G. We conclude that in order to refute the claim that G has an eigenvalue at most β, we must have $G_{n+1}, G_{n+2} \in S'$. This implies that S' contains at most k terms G_i for $i \in [n]$. Then, consider the string x which has ones precisely at these at most k positions $i \in [n]$ corresponding to $G_i \in S'$. It follows that the state $|x, y\rangle_{\text{hist}}$ lies in the null space of all terms in S' with the possible exception of G_{n+2}. Moreover, since V rejects all strings of Hamming weight at most k, there exists by the definition of a cQMA circuit and Lemma 3.19 a $|y\rangle \in \mathcal{B}^{\otimes m}$ such that

$$\text{Tr}\left(G_{n+2}|x, y\rangle\langle x, y|_{\text{hist}}\right) = 1 - \text{Tr}\left(H|x, y\rangle\langle x, y|_{\text{hist}}\right) \leq 1 - b = \beta, \tag{3.19}$$

completing the proof. □

The following two lemmas are required for the proof of Theorem 3.21. Their statements and proofs assume the notation of Theorem 3.21.

Lemma 3.22. *The smallest non-zero eigenvalue of* $Y_2 = \Delta(H_{\text{in}} + H_{\text{prop}} + H_{\text{stab}})$ *scales as* $\Omega(\Delta/L^3)$.

Proof. We bound the smallest non-zero eigenvalue of $H_{\text{in}} + H_{\text{prop}}$; it is straightforward to show using the approach of Reference [171] that the addition of H_{stab} does not affect this lower bound (see Section 1.5.5). Our proof idea here is to "lift" the null space of $H_{\text{in}} + H_{\text{prop}}$ so that the smallest non-zero eigenvalue of $H_{\text{in}} + H_{\text{prop}}$ becomes the smallest eigenvalue of the lifted operator, and then apply the Geometric Lemma (Lemma 1.8) to lower bound the latter.

To begin, recall that the null space of $H_{\text{in}} + H_{\text{prop}}$ consists of all valid history states

$$|\psi\rangle_{\text{hist}} = \frac{1}{\sqrt{L+1}} \sum_{i=0}^{L} V_i \cdots V_1 |\psi\rangle_{A,B} \otimes |0\rangle_C \otimes |i\rangle_D, \tag{3.20}$$

for any $|\psi\rangle_{A,B}$. (Since we omit H_{stab} for now, we assume here that the clock register is represented in binary, i.e. there are no invalid clock states.) As done in Reference [171] and Section 1.5.5, our analysis is simplified by first applying the unitary change of basis $W = \sum_{j=0}^{L} V_1^\dagger \cdots V_j^\dagger \otimes |j\rangle\langle j|$, yielding

$$W|\psi\rangle_{\text{hist}} = |\psi\rangle_{A,B} \otimes |0\rangle_C \otimes |\gamma\rangle_D \tag{3.21}$$

$$WH_{\text{in}}W^\dagger = H_{\text{in}} = I_{A,B} \otimes \left(\sum_{i=1}^{p} |1\rangle\langle 1|_{C_i}\right) \otimes |0\rangle\langle 0|_D \tag{3.22}$$

$$WH_{\text{prop}}W^\dagger = I_{A,B} \otimes I_C \otimes E_D \tag{3.23}$$

where $|\gamma\rangle := \left(\frac{1}{\sqrt{L+1}} \sum_{i=0}^{L} |i\rangle\right)$, and for some operator E_D whose eigenvalues are given by $\lambda_k = 1 - \cos(\pi k/(L+1))$ for $0 \leq k \leq L$ and whose unique zero-eigenvector is $|\gamma\rangle$.

As alluded to above, we now lift the null space of $W(H_{\text{in}} + H_{\text{prop}})W^\dagger$. Letting Π_{hist} denote the projector onto the space of valid history states $|\psi\rangle_{\text{hist}}$, this is accomplished by defining

$$A_1 := W(H_{\text{in}} + p\Pi_{\text{hist}})W^\dagger \tag{3.24}$$
$$A_2 := W(H_{\text{prop}} + 2\Pi_{\text{hist}})W^\dagger. \tag{3.25}$$

Note that $[H_{\text{in}}, \Pi_{\text{hist}}] = [H_{\text{prop}}, \Pi_{\text{hist}}] = 0$, $\| H_{\text{in}} \|_\infty \leq p$ and $\| H_{\text{prop}} \|_\infty \leq 2$. It thus remains to lower bound the smallest eigenvalue of $A_1 + A_2$, for which we apply Lemma 1.8 (Geometric Lemma) to $A_1 + A_2$ via the approach of Reference [171]. For this, we require values for the parameters v and $\alpha(\mathcal{L}_1, \mathcal{L}_2)$.

For v, note that since A_1 is a sum of commuting orthogonal projectors, its smallest non-zero eigenvalue is at least 1 (assuming $p \geq 1$). Similarly, one infers from the spectrum of E_D stated above that the smallest non-zero eigenvalue of A_2 scales as $\Omega(1/L^2)$. It follows that $v \in \Omega(1/L^2)$. As for $\alpha(\mathcal{L}_1, \mathcal{L}_2)$, note that the null spaces $\mathcal{L}_1$ and $\mathcal{L}_2$ can be written as

$$\mathcal{L}_1 = \mathcal{B}_{A,B}^{\otimes(n+m)} \otimes \text{span}(|\psi\rangle \ : \ \langle\psi|0\cdots0\rangle = 0)_C \otimes \text{span}(|1\rangle, \ldots, |L\rangle)_D \oplus \tag{3.26}$$
$$\mathcal{B}_{A,B}^{\otimes(n+m)} \otimes |0\cdots0\rangle_C \otimes \text{span}(|\psi\rangle \ : \ \langle\psi|\gamma\rangle = 0)_D, \tag{3.27}$$
$$\mathcal{L}_2 = \mathcal{B}_{A,B}^{\otimes(n+m)} \otimes \text{span}(|\psi\rangle \ : \ \langle\psi|0\cdots0\rangle = 0)_C \otimes |\gamma\rangle_D. \tag{3.28}$$

Observe that $\mathcal{L}_1 \cap \mathcal{L}_2 = \{\mathbf{0}\}$, as required by Lemma 1.8. Then, letting $\Pi_{\mathcal{L}_1}$ denote the projector onto $\mathcal{L}_1$, we analyze

$$\cos^2 \alpha(\mathcal{L}_1, \mathcal{L}_2) = \max_{\text{unit } |x\rangle \in \mathcal{L}_1, |y\rangle \in \mathcal{L}_2} |\langle x|y\rangle|^2 = \max_{\text{unit } |y\rangle \in \mathcal{L}_2} \langle y|\Pi_{\mathcal{L}_1}|y\rangle = \max_{\text{unit } |y\rangle \in \mathcal{L}_2} \langle y|\Pi_1 + \Pi_2|y\rangle, \tag{3.29}$$

where Π_1 and Π_2 project onto the spaces in Equations (3.26) and (3.27), respectively. As $\langle y|\Pi_2|y\rangle = 0$, we simply need to maximize $\langle y|\Pi_1|y\rangle$, which is equivalent to maximizing $|\langle\psi|\gamma'\rangle|^2$ for any unit vector $|\psi\rangle$ in register D and for unnormalized state $|\gamma'\rangle := (\frac{1}{\sqrt{L+1}} \sum_{i=1}^{L} |i\rangle)$. By the Cauchy-Schwarz inequality, this quantity is upper bounded by $L/(L+1)$. We thus obtain the bound $\cos \alpha(\mathcal{L}_1, \mathcal{L}_2) \leq \sqrt{L/(L+1)}$. Combining this with the identity $2\sin^2 \frac{x}{2} = 1 - \cos x$ and the Maclaurin series expansion for $\sqrt{1+x}$ (where $|x| \leq 1$) yields $2\sin^2 \frac{\alpha(\mathcal{L}_1, \mathcal{L}_2)}{2} \geq \frac{1}{2(L+1)}$. Substituting into Lemma 1.8, the desired result follows. □

Lemma 3.23. *Define* $\Pi_{\text{hist}} := \sum_{x\in\{0,1\}^n, y\in\{0,1\}^m} |x,y\rangle\langle x,y|_{\text{hist}}$ *as the projector onto* $\mathcal{S}_{\text{hist}}$*, let* $\zeta := 2(1+2^{2(n+m)})/(L+1)$*, and consider* $T \subseteq [n]$*. Then, if* V *outputs one with probability at least* $1-\epsilon$ *for inputs* $(x, |y\rangle)$ *with* $x \in \{0,1\}^n$ *such that* $x_i = 1$ *for all* $i \in T$ *and for all* m*-qubit* $|y\rangle$*, one has*

$$\Pi_{\text{hist}} \left[H_{\text{out}} - \sum_{i\in T} G_i \right] \Pi_{\text{hist}} \preceq \zeta\epsilon I. \tag{3.30}$$

Proof. Define $Z_1 := \Pi_{\text{hist}}(-\sum_{i\in T} G_i)\Pi_{\text{hist}}$ and $Z_2 := \Pi_{\text{hist}} H_{\text{out}} \Pi_{\text{hist}}$. Letting $z \in \{0,1\}^n$ denote the characteristic vector of T, i.e. the ith bit of z is set to one if and only if $i \in T$, it follows that any state $|x,y\rangle_{\text{hist}}$ is an eigenvector of Z_1 with eigenvalue $\langle x|z\rangle - |T|$. Hence, for example, $\text{Tr}\left(Z_1 |1^n, y\rangle\langle 1^n, y|_{\text{hist}}\right) = 0$. Further, since V accepts a *non-empty* monotone set, it must accept input $(1^n, |y\rangle)$ with probability at least $1-\epsilon$, implying $\text{Tr}(Z_2 |1^n, y\rangle\langle 1^n, y|_{\text{hist}}) \leq \frac{\epsilon}{L+1}$. This yields an upper bound of

$$\text{Tr}((Z_1 + Z_2)|1^n, y\rangle\langle 1^n, y|_{\text{hist}}) \leq \frac{\epsilon}{L+1} \tag{3.31}$$

in this simple case. We now show that deviating from $|1^n, y\rangle_{\text{hist}}$ above cannot increase our expected value against $Z_1 + Z_2$ by "too much".

To do so, let $|\phi\rangle = \alpha_1|\phi_1\rangle + \alpha_2|\phi_2\rangle$ be an arbitrary valid history state where $|\alpha_1|^2 + |\alpha_2|^2 = 1$, $|\phi_1\rangle$ is a (normalized) superposition of valid history states where each history state in the superposition has a string x in register A at time zero satisfying $x_i = 1$ if $i \in T$, and where $|\phi_2\rangle$ is a valid history state in the space orthogonal to space of all possible states $|\phi_1\rangle$. We thus first have that

$$\text{Tr}\left(Z_1 |\phi\rangle\langle\phi|\right) \leq 0 + \alpha_2^2 \text{Tr}\left(Z_1 |\phi_2\rangle\langle\phi_2|\right) \leq \alpha_2^2[(|T|-1) - |T|] \leq -|\alpha_2|^2. \tag{3.32}$$

Moving on to Z_2, observe that straightforward expansion yields

$$\begin{aligned} \text{Tr}(Z_2|\phi\rangle\langle\phi|) &= |\alpha_1|^2 \text{Tr}(Z_2|\phi_1\rangle\langle\phi_1|) + |\alpha_2|^2 \text{Tr}(Z_2|\phi_2\rangle\langle\phi_2|) & (3.33)\\ &+ \alpha_1\alpha_2^* \text{Tr}(Z_2|\phi_1\rangle\langle\phi_2|) + \alpha_1^*\alpha_2 \text{Tr}(Z_2|\phi_2\rangle\langle\phi_1|). & (3.34) \end{aligned}$$

To upper bound this quantity, we use the fact that $\langle a|b\rangle + \langle b|a\rangle \leq \langle a|a\rangle + \langle b|b\rangle$ for complex vectors $|a\rangle$ and $|b\rangle$. Namely, setting $|a\rangle := \alpha_1\sqrt{Z_2}|\phi_1\rangle$ and $|b\rangle := \alpha_2\sqrt{Z_2}|\phi_2\rangle$ yields

$$\begin{aligned} \text{Tr}(Z_2|\phi\rangle\langle\phi|) &\leq 2|\alpha_1|^2 \text{Tr}(Z_2|\phi_1\rangle\langle\phi_1|) + 2|\alpha_2|^2 \text{Tr}(Z_2|\phi_2\rangle\langle\phi_2|) & (3.35)\\ &\leq 2|\alpha_1|^2 \text{Tr}(Z_2|\phi_1\rangle\langle\phi_1|) + 2|\alpha_2|^2 \frac{1}{L+1}, & (3.36) \end{aligned}$$

where the second inequality follows since $\| Z_2 \|_\infty \leq 1/(L+1)$. Finally, in order to upper bound the term $\mathrm{Tr}(Z_2|\phi_1\rangle\langle\phi_1|)$ in Equation (3.36), observe that since by assumption $\mathrm{Tr}(Z_2|x,y\rangle\langle x,y|_{\mathrm{hist}}) \leq \frac{\epsilon}{L+1}$ for all x with $x_i = 1$ for $i \in T$, and since H_{out} is a projector, it follows that the norm of $H_{\mathrm{out}}|x,y\rangle_{\mathrm{hist}}$ is at most $\sqrt{\epsilon/(L+1)}$. Using the Cauchy-Schwarz inequality, this implies that each cross term in the expansion of $\mathrm{Tr}(Z_2|\phi_1\rangle\langle\phi_1|)$ can contribute a value of magnitude at most $\epsilon/(L+1)$. Since there are at most $2^{2(n+m)}$ such cross terms, and since the non-cross terms are weighted by a convex combination, we hence have the upper bound of $\mathrm{Tr}(Z_2|\phi_1\rangle\langle\phi_1|) \leq (1+2^{2(n+m)})\epsilon/(L+1)$. Combining these bounds, we have

$$\begin{aligned}
\mathrm{Tr}((Z_1+Z_2)|\phi\rangle\langle\phi|) &\leq -|\alpha_2|^2 + \frac{2|\alpha_1|^2(1+2^{2(n+m)})\epsilon}{L+1} + \frac{2|\alpha_2|^2}{L+1} && (3.37)\\
&= \frac{2|\alpha_1|^2(1+2^{2(n+m)})\epsilon + |\alpha_2|^2(1-L)}{L+1} && (3.38)\\
&\leq \frac{2(1+2^{2(n+m)})}{L+1}\epsilon && (3.39)\\
&= \zeta\epsilon && (3.40)
\end{aligned}$$

where the second inequality holds when $L \geq 1$. □

Finally, we show that QIRR is cq-Σ_2-hard to approximate.

Theorem 3.24. *There exists a polynomial time reduction which, given an instance of an arbitrary* cq-Σ_2 *problem* Π*, outputs an instance of QIRR with threshold parameters* h *and* h' *satisfying* $h'/h \in \Theta(N^\epsilon)$ *for some* $\epsilon > 0$*, where* N *is the encoding size of the QIRR instance.*

Proof. We begin by applying Theorems 3.20 and 3.21 to reduce the instance of Π to an instance $(S = \{G_i\}_{i=1}^{n+2}, \alpha, \beta, g, g')$ of QSSC, and henceforth assume the terminology and definitions introduced in Theorem 3.21. Recall that any cover in this QSSC instance must include the terms G_{n+1} and G_{n+2}. For ease of exposition, we first reduce this instance to QIRR with parameters $h = g + 2r - 3$ and $h' = g' + 2r - 3$, where recall r is the number of terms in $H = \sum_{i=1}^r H_i$. This, however, does not suffice to obtain a hardness of approximation gap, as tracing through Theorems 3.20 and 3.21 yields $r \in \omega(g), \omega(g')$, implying $h'/h \to 1$ as the instance Π in Theorem 3.20 grows in size. We then slightly modify our reduction to improve the threshold parameters to $h = gr - 1$ and $h' = g'r - 1$, which yield the desired hardness of approximation gap.

We now state our instance $(T, \gamma, \delta, h, h')$ of QIRR, and follow with an intuitive explanation. For simplicity of exposition, we assume r is a power of two, but our construction can be easily modified to handle the complementary case. We also label $H_r = H_{\text{out}}$. We now introduce three registers: a "tag" qubit register (denoted A), the space the original cover $\mathcal{S}$ acts on (denoted B), and $\log r$ "chaperone" qubits (denoted C). The Hamiltonian terms we define for QIRR, $T := \{F_i\}_{i=1}^{n+2r-1}$, act on $A \otimes B \otimes C = \mathcal{B} \otimes \mathcal{B}^{\otimes(n+m+p+q)} \otimes \mathcal{B}^{\otimes \log r}$, and are defined as:

$$
\begin{aligned}
F_1 &:= |0\rangle\langle 0|_A \otimes (G_1)_B \otimes I_C && (3.41)\\
&\vdots \\
F_n &:= |0\rangle\langle 0|_A \otimes (G_n)_B \otimes I_C && (3.42)\\
F_{n+1} &:= (\Delta+1)\left[|0\rangle\langle 0|_A \otimes (H_1)_B \otimes I_C + |1\rangle\langle 1|_A \otimes I_B \otimes |0\rangle\langle 0|_C\right] && (3.43)\\
&\vdots \\
F_{n+r-1} &:= (\Delta+1)\left[|0\rangle\langle 0|_A \otimes (H_{r-1})_B \otimes I_C + |1\rangle\langle 1|_A \otimes I_B \otimes |r-2\rangle\langle r-2|_C\right] \\
F_{n+r} &:= |0\rangle\langle 0|_A \otimes (I - H_1)_B \otimes I_C + |1\rangle\langle 1|_A \otimes I_B \otimes |r-1\rangle\langle r-1|_C && (3.44)\\
&\vdots \\
F_{n+2r-1} &:= |0\rangle\langle 0|_A \otimes (I - H_r)_B \otimes I_C + |1\rangle\langle 1|_A \otimes I_B \otimes |r-1\rangle\langle r-1|_C. && (3.45)
\end{aligned}
$$

We set $\gamma := \alpha + r - 1$, $\delta := \beta + r - 1$, $h := g + 2r - 3$, and $h' := g' + 2r - 3$. Note that each F_j is a projection up to scalar multiplication, as required. We now provide the intuition behind the construction. QIRR is stated in terms of projectors F_j (up to scalar multiplication), whereas QSSC is stated in terms of Hermitian operators G_i. Hence, in order to move from the latter to the former, a natural idea is to treat each local Hamiltonian term in the sums comprising G_{n+1} and G_{n+2} as distinct terms $F_{n+1}, \ldots, F_{n+r-1}$ and $F_{n+r}, \ldots, F_{n+2r-1}$, respectively. The problem with this approach is that in order to rigorously argue that the gap between thresholds g and g' for QSSC is preserved when defining thresholds h and h' for QIRR, we would like, for example, that *all* terms F_j making up G_{n+1} are chosen together in any candidate cover $T' \subseteq T$. To address this issue, we introduce the chaperone qubits, which ensure that any candidate T' plays by these rules. In particular, we can make sure that all terms $F_{n+1}, \ldots, F_{n+2r-1}$ are chosen in any T', allowing us to rigorously apply our knowledge of the spectra of G_{n+1} and G_{n+2} to the analysis of F_T versus $F_{T'}$.

We now show that if there exists a cover $S' \subseteq S$ for QSSC of size v, then there exists a $T' \subseteq T$ such that $|T'| = v + 2r - 3$ satisfying the conditions for a YES instance of QIRR. Namely, let

$$
T' = \{F_i\}_{i \in [n] \text{ and } G_i \in S'} \cup \{F_{n+1}, \ldots, F_{n+2r-1}\}. \tag{3.46}
$$

Note that it suffices to show that $F_{T'} \succeq \gamma I$ (since if $F_{T'} \succeq \gamma I$, then $F_T \succeq \gamma I$ as well). To show this, observe first that we can write $F_{T'} = K_1 + K_2$, for K_1 and K_2 defined as:

$$\begin{aligned} K_1 &:= |0\rangle\langle 0|_A \otimes \left(\sum_{i \in [n] \text{ and } G_i \in S'} G_i + (\Delta + 1) \sum_{i=1}^{r-1} H_i + \sum_{i=1}^{r} (I - H_i) \right)_B \otimes I_C & (3.47) \\ &= |0\rangle\langle 0|_A \otimes (G_{S'} + (r-1)I)_B \otimes I_C & (3.48) \\ K_2 &:= |1\rangle\langle 1|_A \otimes I_B \otimes \left((\Delta + 1) \left(\sum_{i=0}^{r-2} |i\rangle\langle i| \right) + r|r-1\rangle\langle r-1| \right)_C & (3.49) \\ &= |1\rangle\langle 1|_A \otimes I_B \otimes \left(rI + (\Delta + 1 - r) \sum_{i=0}^{r-2} |i\rangle\langle i| \right)_C , & (3.50) \end{aligned}$$

where we can assume without loss of generality that $\Delta \geq r - 1$. Let $|\phi\rangle = a_0|0\rangle_A|\phi_0\rangle_{BC} + a_1|1\rangle_A|\phi_1\rangle_{BC}$ be an arbitrary state acting on this space with $|a_0|^2 + |a_1|^2 = 1$ and for some unit vectors $|\phi_0\rangle_{BC}$ and $|\phi_1\rangle_{BC}$. Then

$$\begin{aligned} \mathrm{Tr}(F_{T'}|\phi\rangle\langle\phi|) &= \mathrm{Tr}(K_1|\phi\rangle\langle\phi|) + \mathrm{Tr}(K_2|\phi\rangle\langle\phi|) & (3.51) \\ &= |a_0|^2 \mathrm{Tr}(K_1|0\rangle\langle 0| \otimes |\phi_0\rangle\langle\phi_0|) + |a_1|^2 \mathrm{Tr}(K_2|1\rangle\langle 1| \otimes |\phi_1\rangle\langle\phi_1|) & (3.52) \\ &\geq |a_0|^2 (\alpha + r - 1) + |a_1|^2 r & (3.53) \\ &\geq \gamma, & (3.54) \end{aligned}$$

where the first inequality follows since $\mathrm{Tr}(X_{AB} I_A \otimes Y_B) = \mathrm{Tr}(\mathrm{Tr}_A(X_{AB}) Y_B)$ and since $G_{S'}$ is a cover by assumption, and the second inequality since $0 \leq \alpha \leq 1$. We conclude that $H_{T'} \succeq \gamma I$, as desired.

We now prove the other direction, namely that if there does not exist a cover $S' \subseteq S$ for QSSC of size v, then all subsets $T' \subseteq T$ of size $|T'| = v + 2r - 3$ satisfy the conditions for a NO instance of QIRR. To see this, note first that any candidate T' must include the terms F_i for $n + 1 \leq i \leq n + r - 1$. This is because if, for example, $F_{n+1} \notin T'$, then vector $|\phi\rangle := |1\rangle_A|\psi\rangle_B|0\rangle_C$ obtains expected value $\Delta + 1 \geq \gamma$ against F_T, but $|\phi\rangle$ is orthogonal to $F_{T'}$. A similar argument holds for the terms F_i with indices $n + r \leq i \leq n + 2r - 1$, since state $|\phi\rangle := |1\rangle_A|\psi\rangle_B|r-1\rangle_C$ obtains expected value $r \geq \gamma$ against F_T, but obtains value at most $r - 1 \leq \delta$ against $F_{T'}$ if there exists an $i \in [n + r, n + 2r - 1]$ such that $i \notin T'$. Thus, for any candidate T' of size $v + 2r - 3$, this leaves $v - 2$ terms to be chosen from $\{F_1, \ldots, F_n\}$. If we now restrict ourselves to states of the form $|0\rangle_A|\psi\rangle_{BC}$, we find that we are reduced to the same argument in the NO direction of Theorem 3.21 – namely, as S is a cover and any $S' \subseteq S$ of size v is not a cover, there must exist a state $|\phi\rangle := |0\rangle_A|\psi\rangle_{BC}$

such that

$$\mathrm{Tr}(|\phi\rangle\langle\phi|F_T) = \mathrm{Tr}\left[\mathrm{Tr}_C(|\psi\rangle\langle\psi|)(G_S + (r-1)I)\right] \geq \alpha + (r-1) \geq \gamma, \tag{3.55}$$

whereas

$$\mathrm{Tr}(|\phi\rangle\langle\phi|F_{T'}) = \mathrm{Tr}\left[\mathrm{Tr}_C(|\psi\rangle\langle\psi|)(G_{S'} + (r-1)I)\right] \leq \beta + (r-1) = \delta. \tag{3.56}$$

This concludes the reduction from QSSC to QIRR with parameters $h = g + 2r - 3$ and $h' = g' + 2r - 3$.

To obtain improved parameters $h = gr - 1$ and $h' = g'r - 1$, we modify the construction above as follows (intuition to follow): The terms F_i for $n+1 \leq n+2r-1$ from the old construction remain unchanged. For $i \in [n]$, we replace each $F_i := |0\rangle\langle 0|_A \otimes (G_i)_B \otimes I_C$ with the r distinct terms:

$$\begin{aligned}
F_{i,1} &:= |0\rangle\langle 0|_A \otimes (G_i)_B \otimes |0\rangle\langle 0|_C, && (3.57)\\
F_{i,2} &:= |0\rangle\langle 0|_A \otimes (G_i)_B \otimes |1\rangle\langle 1|_C, && (3.58)\\
&\vdots \\
F_{i,r} &:= |0\rangle\langle 0|_A \otimes (G_i)_B \otimes |r-1\rangle\langle r-1|_C. && (3.59)
\end{aligned}$$

Thus, the total number of terms in our QIRR instance increases from $n+2r-1$ to $r(n+2)-1$. Intuitively, we have used the chaperone qubits to split each F_i into r terms $F_{i,j}$, such that if in the old construction we chose $F_i \in T'$, then in the new construction we must place all r terms $F_{i,j}$ in T' in order for the new $F_{T'}$ to maintain its desired spectrum. Thus, whereas the old construction chose $g-2$ terms F_i to place in T', the new construction chooses $r(g-2)$ terms $F_{i,j}$ to place in T', yielding the desired thresholds $h = gr - 1$ and $h' = g'r - 1$.

The completeness and soundness proofs now follow similarly to the previous case. Namely, given a cover $S' \subseteq S$ for QSSC of size v, the set $T' \subseteq T$ with $|T'| = vr - 1$ we choose is

$$T' = \{F_{i,j}\}_{i\in[n] \text{ and } G_i \in S', j\in[r]} \cup \{F_{n+1}, \ldots, F_{n+2r-1}\}. \tag{3.60}$$

Since $F_{T'}$ in this new reduction is precisely $F_{T'}$ in the old reduction, the remainder of this direction proceeds identically. Conversely, if there does not exist a cover $S' \subseteq S$ for QSSC of size v, we similarly first argue that F_i for $n+1 \leq i \leq n+2r-1$ must be chosen in any candidate $T' \subseteq T$ of size $|T'| = vr - 1$, leaving $r(v-2)$ terms to be chosen from $\{F_{1,1}, \ldots, F_{n,r}\}$. This implies that for any such T', there must exist a $j \in [r]$ such that the number of terms $F_{i,j}$ in T' is at most $v-2$. Since no cover of size v exists for our QSSC instance, we conclude there exists an appropriate choice of $|\phi\rangle := |0\rangle_A|\psi\rangle_B|j\rangle_C$ such that Equations (3.55) and (3.56) still hold. □

3.4 Improvements to hardness gaps

We now improve the hardness gaps of Theorems 3.20, 3.21, and 3.24 to obtain the results claimed in Theorems 3.5 and 3.7. The key idea is to use the fact that the gap for QMW from Theorem 3.20 can be amplified by composing the cQMA circuit W with itself. The results here adapt Section 5 of [235] in a simple manner to the quantum setting.

Specifically, assume for the moment that the output qubit of W is actually a classical bit, i.e. that the output qubit is given *after* being measured in the computational basis. Then, one can recursively define $W^1 := W$ and W^t as W^{t-1} with n independent copies of W at each of its n INPUT bits. (Note that entanglement between quantum proofs for different copies of W does not affect the soundness of W^t, as each W outputs a classical bit, and no quantum proofs are reused.) Now, such a recursive composition of W can easily be made well-defined even if W's output qubit is a superposition of $|0\rangle$ and $|1\rangle$ using the principle of deferred measurement [200] – namely, without loss of generality, we can assume W first copies its n classical INPUT bits to an ancilla, and henceforth acts only on its CHOICE and ancilla registers. Thus, the output qubit of each copy of W in W^t is effectively used only as a classical control in the remainder of the circuit, and so the measurement of all output qubits can be deferred to the end of W^t. Finally, since we can assume using standard error reduction that the completeness and soundness error of W scale as 2^{-n}, it follows by the union bound that with probability exponentially close to 1, all the W circuits comprising W^t output the correct answer. In other words, with high probability, one can think of W^t as a composition of zero-error circuits W (where zero-error means zero completeness and soundness error). With this viewpoint, the proof of Lemma 3 of Reference [235] directly yields the following result in the quantum setting.

Lemma 3.25. *If W is a cQMA circuit accepting exactly a monotone set, it follows that:*

1. *$|W^t| \leq n^t |W|$, where $|W|$ denotes the size of W,*
2. *W accepts an input of Hamming weight k if and only if W^t accepts an input of weight k^t,*
3. *W^t accepts exactly a monotone set.*

To improve the hardness gap of Theorem 3.20, we now simply replace the cQMA circuit W constructed in the proof of Theorem 3.20 with W^t for an appropriate choice of t. The details and resulting analysis follow identically to the proof of Theorem 4 of Reference [235], which combined with the improved disperser construction of Reference [232] (see Theorem 7.2 therein) yields:

Theorem 3.26. *QMW is* cq-Σ_2*-hard to approximate with gap* $N^{1-\epsilon}$ *for any* $\epsilon > 0$*, for* N *the encoding size of the QMW instance.*

Using this as the starting point in our reduction chain to QSSC and QIRR, a closer analysis of the proofs of Theorems 3.5 and 3.7 now yields:

Corollary 3.27. *QSSC and QIRR are* cq-Σ_2*-hard to approximate with gaps* $N^{1-\epsilon}$ *and* $N^{\frac{1}{2}-\epsilon}$ *for any* $\epsilon > 0$*, respectively, and where* N *is the encoding size of the respective QSSC and QIRR instances.*

3.5 Hardness of approximation for QCMA

We now briefly remark that the approach of Theorems 3.20 and 3.26 can be adapted to show hardness of approximation for QCMA. Our result is a straightforward extension of Umans' classical result [235] showing NP-hardness of approximation for the problem MONOTONE MINIMUM SATISFYING ASSIGNMENT.

Specifically, define the problem QUANTUM MONOTONE MINIMUM SATISFYING ASSIGNMENT (QMSA) analogously to QMW, except with the definition of a cQMA circuit V modified to drop the second (quantum) proof, i.e. V now only takes one input register comprised of n classical bits. (For example, Definition 3.12 is modified to say that V *accepts* $x \in \{0,1\}^n$ in INPUT if measuring $|a\rangle$ in the computational basis yields 1 with probability at least 2/3.) Then, it is straightforward to re-run the proofs of Theorems 3.20 and 3.26 without the existence of a second quantum proof register, leading to Theorem 3.8.

3.6 A canonical cq-Σ_2-complete problem

In this section, we first show that a quantum generalization of the canonical Σ_2^p-complete problem Σ_2SAT, denoted cq-Σ_2LH, is cq-Σ_2-complete. We then observe that a similar proof yields cq-Σ_2-hardness of approximation for an appropriately defined variant of cq-Σ_2LH.

Definition 3.28 (cq-Σ_2LH)**.** *Given a 3-local Hamiltonian* H *acting on* $N = n + m$ *qubits, and* $a, b \in \mathbb{R}$ *such that* $a \leq b$ *for* $b - a \geq 1$*, output:*

- *YES if* $\exists\ x \in \{0,1\}^n$ *such that* $\forall\ |y\rangle \in \mathcal{B}^{\otimes m}$, $\mathrm{Tr}(H|x\rangle\langle x| \otimes |y\rangle\langle y|) \geq b$.
- *NO if* $\forall\ x \in \{0,1\}^n$, $\exists\ |y\rangle \in \mathcal{B}^{\otimes m}$ *such that* $\mathrm{Tr}(H|x\rangle\langle x| \otimes |y\rangle\langle y|) \leq a$.

Theorem 3.29. cq-Σ_2LH *is* cq-Σ_2*-complete.*

Proof. That cq-Σ_2LH $\in$ cq-Σ_2 follows from Kitaev's verifier for placing k-local Hamiltonian in QMA [171] (see Section 1.5.5). As for cq-Σ_2-hardness, for simplicity we show the result for the case of cq-Σ_2LH defined with 5-local Hamiltonians. The proof for the 3-local case follows identically by instead substituting the 3-local circuit-to-Hamiltonian construction of Reference [164] below (this is possible because our proof does not exploit the structure of the clock register or H_{stab}).

To see that any instance Π of a problem in cq-Σ_2 reduces to an instance of cq-Σ_2LH, let V'' denote the cq-Σ_2 verification circuit for Π. Recall that V'' acts on a classical proof register A, a quantum proof register B, and an ancilla register C. We begin by modifying V'' to obtain a new equivalent circuit V' which first copies the (classical) contents of A to its ancilla register C, and henceforth acts on this copied proof in C throughout the verification. This ensures the contents of A remain unchanged during the verification. Next, we modify V' to obtain V by concatenating to its end a Pauli X on the output qubit; this swaps the cases in which V' accepts and rejects, respectively. This is necessary because if $|c\rangle \otimes |q\rangle$ is accepted by V', then $|c,q\rangle_{\text{hist}}$ obtains low energy against Kitaev's Hamiltonian, whereas in our YES instance here we require high energy. Finally, we apply Kitaev's circuit-to-Hamiltonian construction from Section 3.2 on V to obtain a 5-local Hamiltonian H.

Suppose now that we have a YES instance of Π, i.e. there exists bit string $|c\rangle$ such that for all quantum states $|q\rangle$, the circuit V'' accepts proof $|c\rangle \otimes |q\rangle$ with probability at least $1-\epsilon$ (and hence V rejects $|c\rangle \otimes |q\rangle$ with probability at least $1-\epsilon$). We show that for all $|\psi\rangle_{B,C,D}$, the state $|c\rangle_A \otimes |\psi\rangle_{B,C,D}$ attains expectation value at least b against H, for b from Lemma 3.19. In other words, letting $\Pi_c := (|c\rangle\langle c|_A \otimes I_{B,C,D})$, we claim

$$\langle c| \otimes \langle\psi|H|c\rangle \otimes |\psi\rangle = \langle c| \otimes \langle\psi|\Pi_c H \Pi_c|c\rangle \otimes |\psi\rangle \geq b. \tag{3.61}$$

To see this, observe first that

$$\Pi_c H_{\text{in}} \Pi_c = |c\rangle\langle c|_A \otimes I_B \otimes \left(\sum_{i=1}^{p} |1\rangle\langle 1|_{C_i}\right) \otimes |0\rangle\langle 0|_D =: |c\rangle\langle c|_A \otimes H'_{\text{in}}, \tag{3.62}$$

$$\Pi_c H_{\text{out}} \Pi_c = |c\rangle\langle c|_A \otimes |0\rangle\langle 0|_{B_1} \otimes I_C \otimes |L\rangle\langle L|_D =: |c\rangle\langle c|_A \otimes H'_{\text{out}}, \tag{3.63}$$

$$\Pi_c H_{\text{stab}} \Pi_c = |c\rangle\langle c|_A \otimes I_{B,C} \otimes \sum_{i=1}^{L-1} |01\rangle\langle 01|_{D_i,D_{i+1}} =: |c\rangle\langle c|_A \otimes H'_{\text{stab}}. \tag{3.64}$$

As for $\Pi_c H_{\text{prop}} \Pi_c$, recall that the verification circuit V consists of two phases: The *copy* phase, consisting of n CNOT gates copying the contents of A to C, and the *verification* phase, consisting of the remaining $L - n$ gates of V. In other words, we can write

$$H_{\text{prop}} = \sum_{j=1}^{n} H_j + \sum_{j=n+1}^{L} H_j, \tag{3.65}$$

where $\sum_{j=1}^{n} H_j$ corresponds to the copy phase and $\sum_{j=n+1}^{L} H_j$ to the verification phase. Since during the verification phase, V does not act on A, we have for all $j > n$ that

$$\begin{aligned} \Pi_c H_j \Pi_c &= |c\rangle\langle c|_A \otimes \left[-\frac{1}{2}(V_j)_{B,C} \otimes |j\rangle\langle j-1|_D - \frac{1}{2}(V_j^\dagger)_{B,C} \otimes |j-1\rangle\langle j|_D + \right. && (3.66) \\ &\qquad \left. \frac{1}{2} I_{B,C} \otimes (|j\rangle\langle j| + |j-1\rangle\langle j-1|)_D \right] && (3.67) \\ &=: |c\rangle\langle c|_A \otimes H'_j. && (3.68) \end{aligned}$$

As for the copy phase, let $|i\rangle\langle i| \otimes I$ act on $\mathcal{B} \otimes \mathcal{B}$ for $i \in \{0, 1\}$. Then, observe that

$$(|i\rangle\langle i| \otimes I)\,\text{CNOT}(|i\rangle\langle i| \otimes I) = |i\rangle\langle i| \otimes X^i, \tag{3.69}$$

where X is the Pauli X operator and $X^i = X$ if $i = 1$ and $X^i = I$ otherwise. This implies that for any step $j \leq n$, i.e. where V applies a CNOT gate with qubit A_j as control and C_j as target, and letting c_j denote the jth bit of c, we have

$$\begin{aligned} \Pi_c H_j \Pi_c &= |c\rangle\langle c|_A \otimes \left[-\frac{1}{2} X_{C_j}^{c_j} \otimes |j\rangle\langle j-1|_D - \frac{1}{2} X_{C_j}^{c_j} \otimes |j-1\rangle\langle j|_D + \right. && (3.70) \\ &\qquad \left. \frac{1}{2} I \otimes (|j\rangle\langle j| + |j-1\rangle\langle j-1|)_D \right] && (3.71) \\ &=: |c\rangle\langle c|_A \otimes H'_j(c), && (3.72) \end{aligned}$$

where the notation $H'_j(c)$ means H'_j is a function of c. Letting $H'_{\text{prop}}(c) := \sum_{i=1}^{n} H'_j + \sum_{i=n+1}^{L} H'_j(c)$ and $H(c) := H'_{\text{in}} + H'_{\text{out}} + H'_{\text{stab}} + H'_{\text{prop}}(c)$, we thus have that

$$\langle c| \otimes \langle\psi|H|c\rangle \otimes |\psi\rangle = \langle\psi|H(c)|\psi\rangle. \tag{3.73}$$

It thus suffices to show that $H(c) \succeq bI$.

To see this, we return to the circuit V, and think of V not as accepting classical input c, but rather as corresponding to a set of circuits $\{V_c\}$, where each V_c is just V with c

hard-wired into register A. In particular, at time step $0 \leq j \leq n$, V_c applies X^{c_j} to qubit C_j. Taking this interpretation, we observe that for any string c, plugging V_c into Kitaev's circuit-to-Hamiltonian yields precisely the Hamiltonian $H(c)$. Thus, since by assumption for our particular choice of c, V'' accepts $|c\rangle \otimes |q\rangle$ for all quantum proofs $|q\rangle$, it follows that V_c rejects all $|q\rangle$ with probability at least $1 - \epsilon$. Hence, Lemma 3.19 implies $H(c) \succeq bI$, as desired.

The converse direction proceeds similarly. Namely, suppose we have a NO instance of Π, i.e. for all bit strings $|c\rangle$, there exists a quantum proof $|q\rangle$ such that V'' rejects $|c\rangle \otimes |q\rangle$ with probability at least $1 - \epsilon$. Then, we wish to show that for all c, there exists a $|\psi_c\rangle$ such that $\langle\psi_c|H(c)|\psi_c\rangle \leq a$, for a from Lemma 3.19. To show this, fix an arbitrary c. Since there exists a $|q\rangle$ such that V_c accepts $|q\rangle$ with probability at least $1 - \epsilon$, it follows that the history state $|\psi_c\rangle := \sum_{i=0}^{L} V_i \cdots V_1 |q\rangle_B \otimes |0 \cdots 0\rangle_C \otimes |i\rangle_D$ indeed satisfies

$$\langle\psi_c|H(c)|\psi_c\rangle = \langle\psi_c|H'_{\text{in}} + H'_{\text{out}} + H'_{\text{stab}} + H'_{\text{prop}}(c)|\psi_c\rangle \leq 0 + a + 0 + 0 = a. \tag{3.74}$$

□

Note that the proof of Theorem 3.9 has a special property — the string c fed into the classical proof register of V'' is mapped directly in our reduction to the candidate ground states $|c\rangle|q\rangle$ for 3-local Hamiltonian H. This means, for example, that if there exists a c with the desired properties for a YES instance of our starting cq-Σ_2 problem, then setting $x = c$ in Definition 3.28 yields that the cq-Σ_2LH instance we have mapped to is also a YES instance. It follows that applying the reduction in the proof of Theorem 3.9 to our hard-to-approximate instance of QMW from Theorem 3.26 directly yields Theorem 3.10, i.e. that the following variant of cq-Σ_2LH, which we call cq-Σ_2LH-HW, is cq-Σ_2-hard to approximate. Intuitively, cq-Σ_2LH-HW is defined analogously to cq-Σ_2LH, except that here the goal is to minimize the Hamming weight of x.

Definition 3.30 (cq-Σ_2LH-HW)**.** *Given a 3-local Hamiltonian H acting on $N = n + m$ qubits, $a, b \in \mathbb{R}$ such that $a \leq b$ for $b - a \geq 1$, and integer thresholds $0 \leq g \leq g'$, output:*

- *YES if there exists $x \in \{0,1\}^n$ of Hamming weight at most g such that for all $|y\rangle \in \mathcal{B}^{\otimes m}$, $\text{Tr}(H|x\rangle\langle x| \otimes |y\rangle\langle y|) \geq b$.*
- *NO if for all $x \in \{0,1\}^n$ of Hamming weight at most g', there exists $|y\rangle \in \mathcal{B}^{\otimes m}$ such that $\text{Tr}(H|x\rangle\langle x| \otimes |y\rangle\langle y|) \leq a$.*

Acknowledgements for this chapter. We thank Richard Cleve, Ashwin Nayak, Sarvagya Upadhyay, and John Watrous for interesting discussions, and especially Oded Regev for many helpful insights, including the suggestion to think about a quantum version of PH.

Chapter 4

QMA variants with polynomially many provers

This chapter is based on [111]:

S. Gharibian, J. Sikora, and S. Upadhyay. QMA variants with polynomially many provers. Available at arXiv.org e-Print quant-ph/1108.0617v1, 2011.

In this chapter, we study three variants of multi-prover quantum Merlin-Arthur proof systems. We first show that the class of problems that can be efficiently verified using polynomially many quantum proofs, each of logarithmic-size, is exactly MQA (also known as QCMA), the class of problems which can be efficiently verified via a classical proof and a quantum verifier. We then study the class BellQMA(poly), characterized by a verifier who first applies unentangled, nonadaptive measurements to each of the polynomially many proofs, followed by an arbitrary but efficient quantum verification circuit on the resulting measurement outcomes. We show that if the number of outcomes per nonadaptive measurement is a polynomially-bounded function, then the expressive power of the proof system is exactly QMA. Finally, we study a class equivalent to QMA(m), denoted SepQMA(m), where the verifier's measurement operator corresponding to outcome *accept* is a fully separable operator across the m quantum proofs. Using cone programming duality, we give an alternate proof of a result of Harrow and Montanaro [128] that shows a perfect parallel repetition theorem for SepQMA(m) for any m.

4.1 Introduction and results

The study of classical proof systems has yielded some of the greatest achievements in theoretical computer science, from the Cook-Levin theorem [72, 179], which formally ushered in the age of NP verification systems and the now ubiquitous notion of NP-hardness, to the more modern PCP theorem [30, 29], which led to significant advancements in our understanding of hardness of approximation. A natural generalization of the class NP to the quantum setting is the class quantum Merlin-Arthur (QMA) [171], where a computationally powerful but untrustworthy prover, Merlin, sends a *quantum* proof to convince an efficient *quantum* verifier, Arthur, that a given input string $x \in \{0,1\}^n$ is a YES-instance for a specified promise problem. (See Definition 1.5 for a formal definition of QMA.) It is easy to see that QMA proof systems are at least as powerful as NP, since the ability to process and exchange quantum information does not prevent Arthur from choosing to act classically.

As discussed in Sections 1.5.2 and 1.5.3, much attention has been devoted to QMA over recent years. We now have a number of problems which are *complete* for QMA, with the quantum analogue of classical constraint satisfaction, the physically well-motivated k-local Hamiltonian problem [171], being the canonical QMA-complete problem. Further, QMA is an extremely robust complexity class that satisfies strong error-reduction properties [191]. However, there still remain important open questions. One natural such question, which is the focus of this chapter, is: *How does allowing multiple unentangled provers affect the expressive power of* QMA*?*

Specifically, unlike in the classical setting where allowing multiple proofs, each quantified by a distinct existential quantifier, is trivially equivalent to a single existentially quantified proof, whether the same logic holds in the quantum setting is a highly non-trivial open question due to the quantum phenomenon known as *entanglement* (see Section 1.6.1). Intuitively, entanglement between multiple proofs can be used by cheating provers to correlate their proofs in a way stronger than possible classically. To this end, in this chapter, we are interested in studying the class QMA(poly) [175], a.k.a. quantum Merlin-Arthur proof systems with polynomially many Merlins, where the verifier receives a polynomial number of quantum proofs which are promised to be unentangled with each other. Despite much effort, little is known (more details under *Previous Work* below) about the structural properties of QMA(poly), except for the obvious containments QMA $\subseteq$ QMA(poly) $\subseteq$ NEXP.

Our Results: We show the following three results regarding variants of QMA(poly).

A complete characterization in the logarithmic-size message setting. Let the class $\mathrm{QMA}_{\log}(\mathrm{poly})$ denote the restriction of the class QMA(poly) to the setting where each prover's proof is at most a *logarithmic* number of quantum bits, or *qubits*. We show:

Theorem 4.1. $\mathrm{QMA}_{\log}(\mathrm{poly}) = \mathrm{MQA}$.

Here, recall from Chapter 1 that MQA, also known as QCMA, is defined as QMA except Merlin's proof is a polynomial-size *classical* string. Theorem 4.1 says that if each prover is restricted to sending short quantum proofs, then one can not only do away with multiple provers, but also of the need for *quantum* proofs altogether.

Towards a non-trivial upper bound on BellQMA(poly). One possible approach to the question of $\mathrm{QMA} \stackrel{?}{=} \mathrm{QMA}(\mathrm{poly})$ is to study BellQMA(poly) [52, 10, 64]. BellQMA(poly) is defined analogously to QMA(poly), except that before applying his verification circuit to the polynomially many unentangled quantum proofs, Arthur must measure each proof using a nonadaptive and unentangled (across all proofs) measurement (we call this *Stage 1* of the verification). He then feeds the resulting *classical* outcomes induced by these measurements into an efficient quantum circuit (we call this *Stage 2*), which implements a two-outcome measurement operation corresponding to outcomes *accept* and *reject*.

The significance of BellQMA(poly) here is that if $\mathrm{QMA} \neq \mathrm{BellQMA}(\mathrm{poly})$, then it follows that $\mathrm{QMA} \neq \mathrm{QMA}(\mathrm{poly})$, since $\mathrm{QMA} \subseteq \mathrm{BellQMA}(\mathrm{poly}) \subseteq \mathrm{QMA}(\mathrm{poly})$. To this end, Brandão has shown that for constant m, $\mathrm{QMA} = \mathrm{BellQMA}(m)$ [52]. Where BellQMA(poly) lies, however, remains open. For example, the techniques used to show $\mathrm{QMA}(2) = \mathrm{QMA}(\mathrm{poly})$ [128] do not seem to yield an analogous result $\mathrm{BellQMA}(2) = \mathrm{BellQMA}(\mathrm{poly})$ as they require entangled measurements (i.e. SWAP test measurements) across multiple proofs, which violate the definition of BellQMA.

To make progress on BellQMA(poly), we introduce the class $\mathrm{BellQMA}[r, m]$, which is defined to be $\mathrm{BellQMA}(m)$ with m provers and the additional restriction that in Stage 1 above, the number of outcomes per proof in Arthur's nonadaptive measurements is upper bounded by r. We then show the following:

Theorem 4.2. *For any polynomially bounded functions $r, m : \mathbb{N} \to \mathbb{N}$, it holds that $\mathrm{BellQMA}[r, m] \subseteq \mathrm{QMA}$ (where the containment holds with equality when $r \geq 2$).*

In other words, BellQMA(poly) cannot be used to show that $\mathrm{QMA} \neq \mathrm{QMA}(\mathrm{poly})$ if the verifier in the BellQMA(poly) protocol is restricted to have a polynomially bounded number

of measurement outcomes per proof in Stage 1. We remark that, in general, the number of such measurement outcomes can be exponential in the input length — the restriction that r be a polynomially bounded function is crucial for the proof of Theorem 4.2. For this reason, our result complements, rather than subsumes Brandão's result [52]. In other words, in our notation, Brandão has shown that $\mathrm{BellQMA}[\exp, \mathrm{const}] = \mathrm{QMA}$, and we show $\mathrm{BellQMA}[\mathrm{poly}, \mathrm{poly}] = \mathrm{QMA}$.

Note that we allow the second stage of the BellQMA(poly) verification procedure above to be *quantum*, as per the definition suggested by Chen and Drucker [64], as opposed to *classical*, as studied by Brandão [52]. The conclusion of Theorem 4.2 holds even if the second stage of verification is completely classical.

Finally, it is worth noting that by combining Theorems 4.1 and 4.2, we conclude that in the setting of BellQMA(poly), if MQA $\neq$ QMA, then having the Merlins send logarithmic-size proofs without any restriction on the number of local measurement outcomes of Arthur in Stage 1 has less expressive power than sending polynomial-size proofs but restricting the number of outcomes, even though the number of measurement outcomes in Stage 1 per Merlin in both cases is the same, i.e. polynomial in the input length.

Perfect parallel repetition for SepQMA(m). A key question in designing proof systems is how to improve the completeness and soundness parameters of a verification protocol without increasing the required number of rounds of communication. A natural approach for doing so is to repeat the protocol multiple times in parallel. With QMA, however, this raises the concern that Merlin might try to cheat by entangling his proofs across these parallel runs. If, though, *perfect parallel repetition* holds, it means that for any input string x, if the verification procedure V accepts with probability $p(|x|)$, then if we run V k times in parallel, the probability of accepting in all k runs of V is precisely $p(|x|)^k$. In other words, if perfect parallel repetition holds, there is no incentive for Merlin to cheat — an honest proof which is a product state across all k runs achieves the maximum success probability.

Our final contribution is an alternate proof of a perfect parallel repetition theorem for a class equivalent to QMA(m), namely SepQMA(m). The theorem was first proved in Harrow and Montanaro [128] in connection with an error reduction technique for QMA(poly). However, our proof is significantly different from theirs and uses the cone programming characterization of QMA(poly). Here, SepQMA(m) is defined as QMA(m) with the restriction that Arthur's measurement operator corresponding to acceptance is an unentangled, or *separable*, operator across the m unentangled proofs. We show:

Theorem 4.3. SepQMA(m) *admits perfect parallel repetition.*

Our alternate proof of Theorem 4.3 is significant in that, to the best of our knowledge, it is the first use of duality theory for a cone program *other* than a semidefinite program to establish a parallel repetition result (note that cone programming generalizes semidefinite programming). We remark that semidefinite programs have been previously used to show perfect or strong parallel repetition theorems for various other models of (single or two-prover) quantum interactive proof systems [71, 127, 165], and that the alternate proof of Theorem 4.3 of Harrow and Montanaro is not based on semidefinite programming. Perfect parallel repetition for SepQMA(m) in itself is interesting, as it has been used to show that error reduction is possible for QMA(m) proof systems [128].

Proof ideas and tools: The proof of our first result, Theorem 4.1, is simple, and is an application of the facts that (1) quantum states of a logarithmic number of qubits can be described to within inverse exponential precision using a polynomial number of classical bits, and conversely that (2) given such a classical description, a logarithmic-size quantum state can be efficiently prepared by a quantum circuit. Hence, roughly speaking, one can replace a polynomial number of logarithmic-size quantum proofs with a single polynomial size classical proof, thereby avoiding the danger of a cheating Merlin using entanglement. Although the proof is simple, one cannot hope for a better characterization using other techniques because the reverse containment, i.e. MQA $\subseteq$ $\text{QMA}_{\log}$(poly), also holds using similar ideas.

More technically challenging is our second result, Theorem 4.2. To show the non-trivial direction BellQMA[poly, poly] $\subseteq$ QMA, we simulate an arbitrary BellQMA[poly, poly] protocol by a QMA protocol using the following observation: Although consolidating m quantum proofs into a single quantum proof raises the possibility of cheating using entanglement, if Arthur is also sent an appropriate classical "consistency-check" string, then a dishonest Merlin can be caught with non-negligible probability. Specifically, in our QMA protocol, we ask a single Merlin to send the m quantum proofs of the original BellQMA protocol (denoted by a single state $|\psi\rangle$), accompanied by a "consistency-check" string $\mathbf{p}$ which is a classical description of the probability distributions obtained as the output of Stage 1. One can think of this as having the QMA verifier *delegate* Stage 1 of the BellQMA verification to Merlin. Arthur then performs a consistency check between $|\psi\rangle$ and $\mathbf{p}$ based on the premise that if Merlin is honest, then $\mathbf{p}$ should arise from running Stage 1 of the original verification on $|\psi\rangle$. If this check passes, then Arthur runs Stage 2 of the BellQMA verification on $\mathbf{p}$. If Merlin tries to cheat, however, we show that the check detects this with non-negligible probability, hence achieving the desired containment. Note that the accuracy of the consistency check crucially uses the fact that there are at most polynomially many outcomes to check for each local measurement of Stage 1.

Our last result, Theorem 4.3, is shown using duality theory for cone programs. In particular, we phrase the maximum acceptance probability of a (possibly cheating) prover for the two-fold repetition of a SepQMA(m) verification protocol as a cone program. We then demonstrate a feasible solution for its dual yielding an upper bound on the maximum acceptance probability. The objective value of this dual solution is precisely the product of the optimum values of the two instances of the SepQMA(m) verification protocols. We conclude that one of the optimal strategies of the provers is to be faithful in the following sense: Each prover elects not to entangle his/her two quantum proofs for the two instances of the SepQMA(m) protocol and instead sends a tensor product of optimal proofs for both the instances.

Previous work. The expressive power of multiple Merlins was first studied by Kobayashi, Matsumoto and Yamakami [175], who showed that QMA(2) = QMA(poly) if and only if the class of QMA(2) protocols with completeness c and soundness s (with at least inverse polynomial gap) is exactly equal to QMA(2) protocols with completeness 2/3 and soundness 1/3. Recently, Harrow and Montanaro [128] demonstrated a *product state test*, wherein given two copies of a *pure* quantum state on multiple systems, the test distinguishes between the cases when the quantum state is a *fully* product state across all the systems or *far* from any such state. Using this test, they answered a few important questions regarding QMA(poly). In particular, they showed that

$$\text{QMA}(2) = \text{QMA}(\text{poly}) \tag{4.1}$$

and that error reduction is possible for such proof systems. Prior to their result, the answers to both the questions were known to be affirmative assuming a *weak* version of the Additivity Conjecture [10]. One of the crucial properties of the product state test is that it can be converted into a QMA(2) protocol, where Arthur's measurement operator corresponding to outcome *accept* is a separable operator across the two proofs. Harrow and Montanaro established a perfect parallel repetition theorem for such proof systems, a crucial step in obtaining exponentially small error probabilities.

Blier and Tapp initiated the study of *logarithmic*-size unentangled quantum proofs [48]. They showed that two unentangled quantum proofs suffice to show that a 3-coloring of an input graph exists, implying that NP has *succinct* unentangled quantum proofs. A drawback of their protocol is that although it has *perfect* completeness, its soundness is only inverse polynomially bounded away from 1. Shortly after, Aaronson, Beigi, Drucker, Fefferman and Shor [10] showed that satisfiability of any 3-SAT formula of size n can be proven by $\widetilde{O}(\sqrt{n})$ unentangled quantum proofs of $O(\log n)$ qubits with perfect completeness

and constant soundness (see also [64]). In a subsequent paper [39], Beigi improved directly on Blier and Tapp's result [48] by showing that by sacrificing perfect completeness, one can show that NP has two logarithmic-size quantum proofs with a better gap between completeness and soundness probabilities than in [48] (see also Chiesa and Forbes [65] and Le Gall, Nakagawa, and Nishimura [103] for related improvements which do not sacrifice perfect completeness).

Finally, one of the open questions raised in Reference [10] concerns the power of Arthur's verification procedure. In particular, the paper introduces two different classes of verification procedures, BellQMA and LOCCQMA verification. Roughly speaking, LOCCQMA verification corresponds to Arthur applying a measurement operation that can be implemented by Local Operations and Classical Communication (LOCC) (with respect to the partition induced by the multiple proofs). The authors raised the question of whether BellQMA(poly) = QMA or not. Brandão [52] showed that BellQMA(m) is equal to QMA for constant m. In a recent development, Brandão, Christandl and Yard [53] showed that LOCCQMA(m) is equal to QMA for constant m.

Open problems. A natural open question concerning the results presented in this chapter is the relationship between BellQMA(poly) and QMA. We believe that understanding the complexity of BellQMA protocols will shed new light on the bigger question pertaining to QMA(2) and QMA. Another avenue of interest is to find further applications of the cone programming characterization of multi-prover quantum Merlin-Arthur proof systems. One question concerning the parallel repetition result presented in this chapter is to investigate whether cone programming duality can be used to analyze the product state test in Reference [128]. Finally, it would be interesting to find other classes of QMA(m) protocols that admit a perfect parallel repetition theorem.

Organization of this chapter. We begin in Section 4.2 with background and notation, defining relevant complexity classes in Section 4.2.1, and reviewing cone programming in Section 4.2.2. Theorems 4.1, 4.2, and 4.3 are proved in Sections 4.3, 4.4, and 4.5, respectively.

4.2 Preliminaries

In this section, we state useful lemmas, and discuss relevant complexity classes and cone programming. Throughout the chapter, we use $|x|$ to denote the length of string $x \in$

$\{0,1\}^*$. The standard Hilbert-Schmidt inner product of operators A and B is denoted $\langle A, B\rangle := \mathrm{Tr}(A^\dagger B)$, where $A^\dagger$ denotes the adjoint of A.

First, a useful lemma in this chapter regarding the trace norm (which is a Schatten p-norm with $p = 1$) is the following:

Lemma 4.4 ([244]). *Let $\{\rho_1 \ldots, \rho_k\} \subset \mathrm{D}(\mathcal{X})$ and $\{\sigma_1, \ldots, \sigma_k\} \subset \mathrm{D}(\mathcal{X})$. Then for any Schatten p-norm,*

$$\left\| \bigotimes_{i=1}^{k} \rho_i - \bigotimes_{i=1}^{k} \sigma_i \right\|_p \leq \sum_{i=1}^{k} \| \rho_i - \sigma_i \|_p . \tag{4.2}$$

Next, generalizing Definition 1.122, we say a (possibly unnormalized) operator $A \in \mathrm{Pos}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m)$ is *fully separable* (i.e. unentangled) if it can be written as

$$A = \sum_{i=1}^{k} P_1(i) \otimes \cdots \otimes P_m(i), \tag{4.3}$$

where $P_j(i) \in \mathrm{Pos}(\mathcal{X}_j)$, for every $j \in [m]$ and $i \in [k]$. We denote the cone of fully separable operators as $\mathrm{Sep}(\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_m)$. In the setting of quantum information, one typically also has $\mathrm{Tr}(A) = 1$. It will be useful to note that the set of fully separable density operators is convex, compact, and has non-empty interior since it contains a ball around the normalized identity operator [124, 125, 126].

4.2.1 Relevant complexity classes

We now define the relevant complexity classes specific to this chapter. Recall that a promise problem $A = (A_{\mathrm{yes}}, A_{\mathrm{no}})$ is a partition of the set $\{0,1\}^*$ into three disjoint subsets: the set A_{yes} denotes the set of YES-instances of the problem, the set A_{no} denotes the set of NO-instances of the problem, and $\{0,1\}^* \backslash (A_{\mathrm{yes}} \cup A_{\mathrm{no}})$ is the set of disallowed strings.

We begin by formally generalizing the definition of QMA (see Definition 1.5) to the setting of m unentangled provers.

Definition 4.5 (QMA(m)). *A promise problem $A = (A_{\mathrm{yes}}, A_{\mathrm{no}})$ is in* QMA(m) *if there exist polynomials p, q and a polynomial-time uniform family of quantum circuits $\{Q_n\}$, where Q_n takes as input a string $x \in \Sigma^*$ with $|x| =: n$, quantum proof $|y\rangle \in \mathrm{Sep}\left(\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_{m(n)}\right)$ where $\mathcal{X}_i := (\mathbb{C}^2)^{\otimes p(n)}$ for $i \in [m(n)]$, and $q(n)$ ancilla qubits in state $|0\rangle^{\otimes q(n)}$, such that:*

- *(Completeness) If* $x \in A_{\text{yes}}$*, then there exists a proof* $|y\rangle \in \text{Sep}\left(\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_{m(n)}\right)$ *such that* Q_n *accepts* $(x, |y\rangle)$ *with probability at least* $2/3$*.*
- *(Soundness) If* $x \in A_{\text{no}}$*, then for all proofs* $|y\rangle \in \text{Sep}\left(\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_{m(n)}\right)$*,* Q_n *accepts* $(x, |y\rangle)$ *with probability at most* $1/3$*.*

The class QMA(poly) *is defined as* $\text{QMA(poly)} := \bigcup_{m \in \text{poly}} \text{QMA}(m)$*.*

For clarity, note that $|y\rangle \in \text{Sep}\left(\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_{m(n)}\right)$ must have the form $|y\rangle = |y_1\rangle \otimes \cdots \otimes \left|y_{m(n)}\right\rangle$ for $|y_i\rangle \in \mathcal{X}_i$. Hence, QMA$(m)$ can be thought of as having $m(n)$ unentangled provers. Note that like QMA = QMA(1), the constants $2/3$ and $1/3$ above can be amplified to values exponentially close to 1 and 0, respectively, by having the verifier run the verification procedure polynomially times in parallel (this requires increasing the number of provers, however). Also, we will use the fact that corresponding to any QMA(m) protocol is a two-outcome POVM (see Section 1.4.2) consisting of operators $\{C_{\text{accept}}, C_{\text{reject}}\}$, such that for any candidate proof $|\psi\rangle = |\psi_1\rangle \otimes \cdots \otimes \left|\psi_{m(n)}\right\rangle$, the probability of the verifier accepting (rejecting) is given by $\text{Tr}(C_{\text{accept}}|\psi\rangle\langle\psi|)$ ($\text{Tr}(C_{\text{reject}}|\psi\rangle\langle\psi|)$).

All complexity classes considered in this chapter are variants of QMA(m) and satisfy the properties mentioned above in Definition 4.5. The next two classes we define are:

1. **[QMA$_{\log}$(poly)]** A subclass of QMA(poly) in which each Merlin's message to Arthur is $O(\log(|x|))$ qubits in length for input string x.
2. **[SepQMA(poly)]** A subclass of QMA(poly), wherein Arthur's measurement operator C_{accept} corresponding to outcome *accept* is a fully separable operator across the proofs.

For clarity, we next give a more formal definition of the variant of BellQMA we introduce, BellQMA$[r, m]$.

Definition 4.6 (BellQMA$[r, m]$)**.** *Let* $r, m : \mathbb{N} \to \mathbb{N}$ *be two functions. A promise problem* $A = (A_{yes}, A_{no})$ *is in class* BellQMA$[r, m]$ *if there exists a* QMA(m) *verification protocol in which Arthur is restricted to act as follows.*

1. *Arthur performs a polynomial-time quantum computation on the input* x *and generates a description of quantum circuits* $V_1(x), \ldots, V_m(x)$*, one for each of the* m *provers.*
2. *(Stage 1) Arthur simultaneously measures all* m *quantum proofs by applying* $V_i(x)$ *to the* i*-th quantum proof, where the action of* $V_i(x)$ *can be described by a unitary operator followed by measurement in the standard basis. The label of the* i*-th measurement outcome is stored as a classical string* y_i *also identified as an element of* $[r(|x|)]$*.*

3. *(Stage 2) Arthur runs an efficient quantum verification circuit on input x and measurement outcomes $(y_1, \ldots, y_m)$ to decide whether to accept or reject.*

Note that the key distinction between BellQMA$[r, m]$ and BellQMA(poly) is that the former has the number of measurement outcomes in Stage 1 of the protocol bounded by $r(|x|)$, whereas the latter may allow exponentially many possible outcomes. Throughout this chapter, we use the notation BellQMA[poly, poly] to denote

$$\text{BellQMA}[\text{poly}, \text{poly}] := \bigcup_{r \in \text{poly}} \bigcup_{m \in \text{poly}} \text{BellQMA}[r, m]. \tag{4.4}$$

4.2.2 Cone programming

We now briefly review basic notions in conic optimization (or cone programming), which is a generalization of semidefinite optimization. The reader is referred to the text of Boyd and Vandenberghe [51] for further details.

To begin, recall that a set K in an underlying Euclidean space is a *cone* if $x \in K$ implies that $\lambda x \in K$ for all $\lambda \geq 0$. A cone K is *convex* if $x, y \in K$ implies that $x + y \in K$. *Cone programs* are concerned with optimizing a linear function over the intersection of a convex cone and an affine space. It generalizes several well-studied models of optimization including semidefinite programming (where $K = \text{Pos}(\mathcal{X})$) and linear programming (where $K = \mathbb{R}^n_+$). In this chapter, we are primarily concerned with the cone of fully separable operators $K = \text{Sep}\,(\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_m)$, which as stated in Section 4.2 is a closed, convex cone with non-empty interior.

A cone program associates the following 4-tuple $(C, b, \mathcal{A}, K)$ to an optimization problem, which we denote as the *Primal* problem:

Primal problem (P)

$$\begin{aligned} \text{supremum:} \quad & \langle X, C \rangle && (4.5)\\ \text{subject to:} \quad & \mathcal{A}(X) = b, && (4.6)\\ & X \in K, && (4.7) \end{aligned}$$

where $\mathcal{A} : \text{Span}(K) \to \mathbb{R}^m$ is a linear transformation, and K lies in a real Euclidean space. (Note that the choice of inner product in $\langle X, C \rangle$ depends on the Euclidean space K lies in.) We say that the cone program is *feasible* if $\{X : \mathcal{A}(X) = b\} \cap K$ is non-empty and

strictly feasible if $\{X : \mathcal{A}(X) = b\} \cap \text{int}(K)$ is non-empty, where $\text{int}(\cdot)$ denotes the interior of a set.

Next, associated with a cone K is its dual cone K^*, defined as

$$K^* = \{S : \langle X, S\rangle \geq 0 \text{ for all } X \in K\}. \tag{4.8}$$

Via the dual cone, for every Primal problem, one can define an associated *Dual* problem as follows:

Dual problem (D)

$$\begin{aligned} \text{infimum:} \quad & \langle b, y\rangle && (4.9)\\ \text{subject to:} \quad & \mathcal{A}^*(y) = C + S, && (4.10)\\ & S \in K^*, && (4.11) \end{aligned}$$

where $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$. We remark that so long as K is closed (which is the case for the cone of fully separable operators), the roles of the Primal and Dual problems can be freely interchanged, since a convex cone K is closed if and only if $K = K^{**}$.

The problems (P) and (D) obey the following special relationship.

Lemma 4.7 (Weak Duality)**.** *If X is primal feasible and (y, S) is dual feasible then*

$$\langle b, y\rangle - \langle X, C\rangle = \langle X, S\rangle \geq 0. \tag{4.12}$$

In other words, let the optimal values of (P) and (D) be denoted p^* and d^*, respectively. Then $p^* \leq d^*$. This raises the important question: Does $p^* = d^*$? In general, this is not the case. However, if indeed $p^* = d^*$, we say that *strong duality* holds. Below we give a condition which, if satisfied, guarantees that strong duality holds.

Theorem 4.8 (Strong Duality)**.** *If* (P) *is strictly feasible, then strong duality holds, i.e. $p^* = d^*$. In particular, this implies that if p^* is finite, then both (P) and* (D) *attain their optimal values, which coincide.*

Note that when K is a closed, convex cone, one can flip the roles of primal and dual problems in Theorem 4.8.

4.3 Equivalence of MQA and $\mathbf{QMA_{log}(poly)}$

We now prove Theorem 4.1, i.e. that MQA = $\mathrm{QMA}_{\log}(\mathrm{poly})$. We first show the direction MQA $\subseteq$ $\mathrm{QMA}_{\log}(\mathrm{poly})$. Let $A = (A_{\mathrm{yes}}, A_{\mathrm{no}})$ be a promise problem in MQA and let $x \in \{0,1\}^n$ be the input string. Suppose the MQA prover sends an m-bit classical proof to the verifier, for polynomially bounded m. Then the following straightforward $\mathrm{QMA}_{\log}(m)$ protocol achieves the desired containment:

1. **Embed classical bits into qubits.** Each (unentangled) prover $i \in [m]$ sends a single qubit $|\psi_i\rangle \in \mathbb{C}^2$ to Arthur. If the i-th prover is honest, his/her qubit is the computational basis state corresponding to the i-th bit of the classical MQA proof.
2. **Make things classical again.** Arthur measures all proofs in the computational basis, obtaining a classical string $y \in \{0,1\}^m$.
3. **Run MQA verification.** Arthur runs the MQA verification circuit on x and y and accepts if and only if acceptance occurs in the MQA verification.

The completeness property follows straightforwardly. The soundness property is also easy to observe. Note that Arthur runs the MQA verification on a classical string y and hence he accepts the string with probability at most 1/3.

To show the reverse containment, let $A = (A_{\mathrm{yes}}, A_{\mathrm{no}})$ be a promise problem in class $\mathrm{QMA}_{\log}(\mathrm{poly})$ and let $x \in \{0,1\}^n$ be the input string. Suppose we have a $\mathrm{QMA}_{\log}(m)$ protocol for polynomially bounded m, where prover i sends a $\lceil c \log n \rceil$-qubit state $|\psi_i\rangle$ for some constant $c > 0$. Let $r(n) = 2^{\lceil c \log n \rceil} = O(n^c)$. The MQA protocol proceeds as follows:

1. **Describe proofs classically.** The prover sends m classical registers represented by the tuple $(\mathsf{C}_1, \mathsf{C}_2, \ldots, \mathsf{C}_m)$, each of length $2n \cdot r(n)$ to Arthur. If the prover is honest, register C_i contains a classical description of the i-th quantum proof $|\psi_i\rangle$ of the $\mathrm{QMA}_{\log}(m)$ protocol.
2. **State preparation.** Using the contents of register C_i, for every choice of $i \in [m]$, Arthur prepares the state $|\psi_i\rangle$ by first determining a unitary U_i such that $U_i|0\ldots0\rangle = |\psi_i\rangle$, and then implementing U_i with high precision using a finite set of approximately universal gates, obtaining states $|\psi_i'\rangle$.
3. **Run $\mathbf{QMA_{log}}(m)$ verification.** Arthur runs the $\mathrm{QMA}_{\log}(m)$ verification circuit on $|\psi_1'\rangle \otimes \cdots \otimes |\psi_m'\rangle$ and accepts if and only if acceptance occurs in $\mathrm{QMA}_{\log}(m)$ verification.

Observe that each classical register C_i is of size polynomial in n, implying the overall proof length is of polynomial size. In Step 1, the prover uses n bits to represent the real and imaginary parts of each of the polynomially many entities ($r(n)$ entries) required to describe each $|\psi_i\rangle$. Let the unit vector described by register C_i be denoted $|\psi_i\rangle$. In Step 2, U_i is easily found, as the unitary that maps $|0\dots0\rangle$ to $|\psi_i\rangle$ is the inverse of the unitary that maps $|\psi_i\rangle$ to $|0\dots0\rangle$. Next, U_i can be efficiently decomposed into a product of U_i' one- and two-qubit unitary gates (see Bernstein and Vazirani [46] for details, or Section 1.5.1 under "Universal gate sets" for a brief discussion) such that $\| U_i - U_i' \|_\infty$ is inverse exponentially small. Since Steps 1 and 2 can be performed to within inverse exponential error, we thus can ensure $\| |\psi_i\rangle - |\psi_i'\rangle \| \leq \epsilon$ for all $i \in [m]$ and for inverse exponential $\epsilon > 0$. By Lemma 4.4, it follows that the overall precision error is at most $m\epsilon$ for polynomial m, and thus the completeness and soundness of the protocol are bounded from below and above by $\frac{2}{3} - m\epsilon$ and $\frac{1}{3} + m\epsilon$, respectively.

4.4 Equivalence of BellQMA[poly, poly] and QMA

We now show Theorem 4.2, i.e. that BellQMA$[r, m]$ = QMA for polynomially-bounded functions r and m. For notational convenience, let $\Pi_j(i)$ denote Arthur's i-th POVM element in Stage 1 of the BellQMA verification protocol for the j-th prover (i.e. $\sum_{i=1}^{r} \Pi_j(i) = I$), where we assume without loss of generality that the number of possible outcomes is exactly r for each prover, and where $j \in [m]$ for m the number of provers.

We proceed as follows. Let $A = (A_{\text{yes}}, A_{\text{no}})$ be a promise problem, and x be an input string of length $n := |x|$. Note first that the containment QMA $\subseteq$ BellQMA[poly, poly] follows since, by definition, QMA $\subseteq$ BellQMA$[2, 1]$. For the reverse containment, suppose we have a BellQMA$[r, m]$ protocol for polynomially bounded functions $r, m : \mathbb{N} \to \mathbb{N}$ with completeness $2/3$ and soundness $1/3$. We show that this protocol can be simulated by a QMA protocol as follows.

Merlin's proof consists of two registers (X, Y), which should be thought of as the *classical* and *quantum* registers, respectively. Suppose optimal proofs for the BellQMA$[r, m]$ protocol for input x are given by ρ_j for $j \in [m]$. Then, in the quantum register Y, an honest Merlin should send many copies of the state ρ_j. Specifically, Y is partitioned into m registers Y_j, one for each original prover, and each Y_j should contain k copies of ρ_j, for k a carefully chosen polynomial. In other words, Y should contain the state $[\rho_1^{\otimes k}]_{\mathsf{Y}_1} \otimes \cdots \otimes [\rho_m^{\otimes k}]_{\mathsf{Y}_m}$. We further view each Y_j as a block of registers $(\mathsf{Y}_j^1, \dots, \mathsf{Y}_j^k)$ where Y_j^l should contain the l-th copy of ρ_j.

In the classical register X, Merlin sends the classical "consistency check" string alluded to in Section 4.1. Specifically, an honest Merlin prepares a quantum state in the computational basis, which intuitively corresponds to a bit string describing the m classical probability distributions Arthur induces upon applying the measurement operation corresponding to Stage 1 of the BellQMA verification to each of the optimal proofs ρ_j, respectively. More formally, we partition X into mr registers X_j^i corresponding to each of the $j \in [m]$ provers and $i \in [r]$ POVM outcomes per prover. The content of X_j^i should be $p_j(i) := \langle \Pi_j(i), \rho_j \rangle$, truncated to α bits of precision (α polynomially bounded), such that $\sum_{i=1}^r p_j(i) = 1$. For example, if the j-th prover's proof was the single qubit state $\rho_j = |0\rangle\langle 0|$, with $\Pi_j(1) = |0\rangle\langle 0|$ and $\Pi_j(2) = |1\rangle\langle 1|$, then $\mathsf{X}_j = (1, 0)$.

Of course, Merlin may elect to be dishonest and choose not to send a proof of the above form to Arthur by, e.g., sending a quantum state which is entangled across the registers (X, Y). To catch this, our QMA protocol is defined as follows:

1. Merlin sends Arthur a quantum state in registers (X, Y), for X and Y defined as above.
2. **Force X to be classical.** Arthur measures register X in the computational basis and reads the measurement outcome. This forces X to essentially be a classical register of bits, and destroys any entanglement or correlations between X and Y.
3. **X should contain probability distributions.** Arthur checks whether the content of registers X_j form a probability distribution p_j. Arthur rejects if this is not the case.
4. **Consistency check: Can the quantum states in Y reproduce the distributions in X?** Arthur picks independently and uniformly at random, an index $j \in [m]$ and another index $i \in [r]$. He applies the measurement $\{\Pi_j(i)\}_{i=1}^r$ separately to each register $\mathsf{Y}_j^1, \ldots, \mathsf{Y}_j^k$, and counts the number of times outcome i appears, which we denote henceforth as $n_j(i)$. Arthur rejects if
$$\left| \frac{n_j(i)}{k} - p_j(i) \right| \geq \frac{1}{p}, \tag{4.13}$$
for p a carefully chosen polynomial.
5. **Run Stage 2 of the BellQMA verification and repeat for error reduction.** For each prover j, Arthur samples an outcome from $[r]$ according to the distribution in $(\mathsf{X}_j^1, \ldots, \mathsf{X}_j^r)$, and runs Stage 2 of the BellQMA verification on the resulting set of samples. He repeats this process independently a polynomial number of times q, and accepts if and only if the BellQMA procedure accepts on the majority of the runs.

Let us discuss the intuition behind the verification procedure above. The key is Step 4, where Arthur cross-checks that the classical distributions sent in X really can be obtained by measuring m quantum proofs, which for an honest Merlin should be unentangled. In this sense, our protocol can alternatively be viewed as using *quantum* proofs (Y) to check validity of a *classical* proof (X). Intuitively, the reason why entanglement in Y does not help a dishonest Merlin in Step 3 is due to the local nature of Arthur's checks/measurements. Finally, once Arthur is satisfied that X contains valid distributions, he runs Step 5. We remark that repetition is used here in order to boost the probability of acceptance in the $x \in A_{\text{yes}}$ case to exponentially close to 1, which is required to separate it from the $x \in A_{\text{no}}$ case, where the probability of catching a dishonest Merlin is only inverse polynomially bounded away from 1. Once such a gap exists, standard error reduction techniques [172, 191] (see Section 1.5.2) can be used to further improve completeness and soundness parameters.

To formally analyze completeness and soundness of the QMA protocol, we assign the following values to the parameters, all of which are polynomial in n in our setting:

$$q = 50n \qquad \text{and} \qquad p = 20mr \qquad \text{and} \qquad k = 5p^3 \qquad \text{and} \qquad \alpha = 20nmr. \tag{4.14}$$

Completeness. Intuitively, when $x \in A_{\text{yes}}$, Merlin passes Step 4 with probability exponentially close to 1 since he has no incentive to cheat — he can send an unentangled proof in Step 1 to Arthur corresponding to the optimal proofs ρ_j in the BellQMA protocol, such that the expected value of $n_j(i)/k$ is indeed $p_j(i)$. Arthur's checks in Step 4 are then independent local trials, allowing a Chernoff bound to be applied. We then show that Merlin passes each run in Step 5 with constant probability, and applying the Chernoff bound a second time yields the desired completeness exponentially close to 1 for the protocol.

To state this formally, suppose Merlin is honest and sends registers (X, Y) in the desired form, i.e., X_j^i contains $p_j(i) = \langle \Pi_j(i), \rho_j \rangle$ up to α bits of precision, and Y_j^l contains ρ_j. Then, the expected value of the random variable $n_j(i)$ is $\mathbb{E}[n_j(i)] = k \langle \Pi_j(i), \rho_j \rangle$, which is equal to $k \cdot p_j(i)$ up to the error incurred by representing $p_j(i)$ using α bits of precision. In other words,

$$\left| \frac{\mathbb{E}[n_j(i)]}{k} - p_j(i) \right| < \frac{1}{2^\alpha} < \frac{1}{2p}. \tag{4.15}$$

We can hence upper bound the probability of rejecting in Step 3 by

$$\Pr\left[\left| \frac{n_j(i)}{k} - p_j(i) \right| \geq \frac{1}{p} \right] < \Pr\left[\left| \frac{n_j(i)}{k} - \frac{\mathbb{E}[n_j(i)]}{k} \right| \geq \frac{1}{2p} \right] \leq 2 \exp\left(-\frac{5p}{4} \right), \tag{4.16}$$

where the first inequality follows from Equation (4.15) and the second from the Chernoff bound. Thus, Merlin passes Step 4 with probability exponentially close to 1.

We now turn to the final step. Since $x \in A_{\text{yes}}$, we know that the optimal distributions, denoted $q_j := (\langle \Pi_j(1), \rho_j \rangle, \ldots, \langle \Pi_j(r), \rho_j \rangle)$ for $j \in [m]$, obtained in Stage 1 of the original BellQMA protocol are now accepted in Stage 2 with probability at least 2/3. However, in our case, Merlin was only able to specify each q_j up to α bits of precision per entry as the distributions p_j. To analyze how this affects the probability of acceptance, let P_j and Q_j be diagonal operators with entries $P_j(i,i) = p_j(i)$ and $Q_j(i,i) = \langle \Pi_j(i), \rho_j \rangle$, respectively. Letting C_{accept} denote the POVM element corresponding to outcome *accept* in Stage 2 of the BellQMA protocol, we thus bound the change in acceptance probability by:

$$\begin{aligned}
\left| \text{Tr} \left[C_{\text{accept}} \left(\bigotimes_{j=1}^{m} P_j - \bigotimes_{j=1}^{m} Q_j \right) \right] \right| &\leq \left\| \bigotimes_{j=1}^{m} P_j - \bigotimes_{j=1}^{m} Q_j \right\|_{\text{tr}} && (4.17) \\
&\leq \sum_{j=1}^{m} \| P_j - Q_j \|_{\text{tr}} && (4.18) \\
&= \sum_{j=1}^{m} \sum_{i=1}^{r} |p_j(i) - \langle \Pi_j(i), \rho_j \rangle| && (4.19) \\
&\leq \frac{mr}{2^{20nmr}}, && (4.20)
\end{aligned}$$

where the first inequality follows from the fact that $|\text{Tr}(AB)| \leq \| A \|_{\infty} \cdot \| B \|_{\text{tr}}$ and the second inequality follows from Lemma 4.4. Therefore, the probability of success for each of the q runs of the BellQMA protocol in Step 5 is at least

$$\left(\frac{2}{3} - \frac{mr}{2^{20nmr}} \right) > 0.6. \tag{4.21}$$

Since each run is independent, applying the Chernoff bound yields that Arthur accepts Merlin's proof in Step 5 with probability at least $1 - 2\exp(-0.02q)$, as desired. There may be some error incurred in sampling, which can be assumed to be exponentially small so that the success probability of each run is still at least 0.6.

Soundness. We now prove that when $x \in A_{\text{no}}$, a dishonest Merlin can win with probability at most inverse polynomially bounded away from 1. To show this, we bound the probability of passing Step 4 by relating the quantity $p_j(i)$ to the expected value of $n_j(i)/k$, and then apply the Markov bound. The desired relationship follows by observing first that the expected value of $n_j(i)/k$ is precisely the probability of obtaining outcome i when measuring proof j of some (honest) unentangled strategy, followed by arguing that the distribution p_j must hence be far from this latter (honest) distribution if Merlin is to pass

Step 5 with probability at least 1/2 (since $x \in A_{\text{no}}$). Combining these facts, we find that Arthur detects a cheating Merlin with inverse polynomial probability in Step 4.

More formally, let the quantum register Y_j contain an arbitrary quantum state σ_j whose reduced states in registers Y_j^l for $l \in [k]$ are given by $\sigma_j(l)$, and define

$$\xi_j := \frac{1}{k} \sum_{l=1}^{k} \sigma_j(l). \tag{4.22}$$

By the linearity of expectation, the expected value of the random variable $n_j(i)/k$ is

$$\mathbb{E}\left[\frac{n_j(i)}{k}\right] = \frac{1}{k} \sum_{l=1}^{k} \langle \Pi_j(i), \sigma_j(l) \rangle = \langle \Pi_j(i), \xi_j \rangle . \tag{4.23}$$

Our goal is to lower bound the expression

$$\Pr\left[\left| \frac{n_j(i)}{k} - p_j(i) \right| \geq \frac{1}{p}\right]. \tag{4.24}$$

To achieve this, we first substitute $p_j(i)$ above with a quantity involving $\mathbb{E}[n_j(i)/k]$, and then apply the Markov bound.

To relate $\mathbb{E}[n_j(i)/k]$ to $p_j(i)$, we first remark that in order for Merlin to pass each run of Step 5 with probability exponentially close to 1, he must send probability distributions p_j, which are accepted by Stage 2 of the BellQMA verification with probability at least 1/2. Let

$$q_j(i) := \langle \Pi_j(i), \xi_j \rangle . \tag{4.25}$$

Let us imagine a BellQMA protocol where the j-th Merlin sends ξ_j as his quantum proof. Since $x \in A_{\text{no}}$, by the soundness property of the BellQMA(m) proof system, the success probability of the Merlins is at most 1/3. In other words, sampling outcomes from the probability distributions $(q_j(1), \ldots, q_j(r))$ and then running the second stage of the BellQMA verification will yield outcome *accept* with probability at most 1/3. Also, observe that

$$\mathbb{E}\left[\frac{n_j(i)}{k}\right] = q_j(i). \tag{4.26}$$

It follows that by letting P_j and Q_j be diagonal operators with the probability vectors p_j and q_j on their diagonals, respectively, and C_{accept} the POVM element corresponding to outcome *accept* in Stage 2 of the BellQMA protocol, we have

$$\frac{1}{10} < \left| \text{Tr}\left[C_{\text{accept}} \left(\bigotimes_{j=1}^{m} P_j - \bigotimes_{j=1}^{m} Q_j \right) \right] \right| \leq \left\| \bigotimes_{j=1}^{m} P_j - \bigotimes_{j=1}^{m} Q_j \right\|_{\text{tr}} \leq \sum_{j=1}^{m} \| P_j - Q_j \|_{\text{tr}} . \tag{4.27}$$

Here, the (loose) lower bound of $1/10$ comes from the following two observations. First, the distributions represented by the Q_j's are derived from a BellQMA protocol and therefore achieve a success probability at most $1/3$ by the soundness property of the BellQMA verification. Second, the distributions represented by the P_j's have to achieve a success probability strictly greater than $1/2$ per run to guarantee that Merlin wins Step 5 with probability exponentially close to 1. Combining these two, we get that the difference between the success probabilities obtained by distributions $\{P_j\}$ and $\{Q_j\}$ should be at least $1/6$ modulo the error incurred due to finite precision when encoding the distributions p_j. The use of the constant $1/10$ overcompensates for this precision error. Hence, there exists a j such that

$$\| P_j - Q_j \|_{\mathrm{tr}} = \sum_{i=1}^{r} |p_j(i) - q_j(i)| \geq \frac{1}{10m}, \tag{4.28}$$

implying the existence of an i such that

$$|p_j(i) - q_j(i)| \geq \frac{1}{10mr}. \tag{4.29}$$

This is our desired relationship between $p_j(i)$ and $\mathbb{E}[n_j(i)/k] = q_j(i)$. Note that the probability of picking pair (i, j) in Step 4 is $1/mr$.

We now substitute this relationship into Equation (4.24) and apply the Markov bound. Specifically, choose i and j as in Equation (4.29), and assume that $p_j(i) > \langle \Pi_j(i), \xi_j \rangle$. Then, we have

$$\Pr\left[\left|\frac{n_j(i)}{k} - p_j(i)\right| < \frac{1}{p}\right] < \Pr\left[\frac{n_j(i)}{k} - \mathbb{E}\left[\frac{n_j(i)}{k}\right] > \frac{1}{10mr} - \frac{1}{p}\right] \leq 1 - \frac{1}{2p}. \tag{4.30}$$

The case of $p_j(i) < \langle \Pi_j(i), \xi_j \rangle$ is similar. We conclude that a dishonest Merlin is caught in Step 4 with probability at least $1/2p$. Therefore, the probability that Arthur proceeds to Step 5 is upper bounded by

$$\left(\frac{1}{mr}\right)\left(1 - \frac{1}{20mr}\right) + \left(1 - \frac{1}{mr}\right)(1) = 1 - \frac{1}{20m^2r^2}, \tag{4.31}$$

where the first term represents the case where Arthur selects the correct pair (i, j) to check, and the second term the complementary case, in which we assume the cheating prover can win with probability 1. Hence the overall success probability of Merlin is at most $1 - 1/20m^2r^2$.

Finally, as mentioned before, since m and r are polynomially bounded functions, we have that the completeness is exponentially close to 1, while the soundness is bounded away

from 1 by an inverse polynomial. By known error reduction techniques for QMA protocols [172, 191], one can amplify the completeness and soundness errors to be exponentially close to 0. This proves our desired containment.

4.5 Perfect parallel repetition for SepQMA(poly)

Using cone programming, we now show Theorem 4.3, i.e., that the class SepQMA(m) admits perfect parallel repetition. Recall now that for C the measurement operator corresponding to outcome *accept*, the maximum success probability of the Merlins in any QMA(m) protocol can be written as the maximum of $\langle \rho, C \rangle$, where ρ is a density operator in the cone $\mathrm{Sep}\left(\mathcal{X}_1, \ldots, \mathcal{X}_m\right)$. This is a simple cone program and can be written as the following primal-dual pair:

Primal problem (P)		Dual problem (D)	
max	$\langle \rho, C \rangle$	min	t
s. t.	$\mathrm{Tr}(\rho) = 1,$	s. t.	$tI_{\mathcal{X}} = C + W,$
	$\rho \in \mathrm{Sep}\left(\mathcal{X}_1, \ldots, \mathcal{X}_m\right),$		$W \in \mathrm{Sep}(\mathcal{X}_1, \ldots, \mathcal{X}_m)^*,$

where $\mathcal{X}$ denotes $\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m$, and $\mathrm{Sep}(\mathcal{X}_1, \ldots, \mathcal{X}_m)^*$ is the dual cone defined as

$$\mathrm{Sep}(\mathcal{X}_1, \ldots, \mathcal{X}_m)^* := \{W : \langle \rho, W \rangle \geq 0 \text{ for all } \rho \in \mathrm{Sep}\left(\mathcal{X}_1, \ldots, \mathcal{X}_m\right)\}. \tag{4.32}$$

(Note that $\mathrm{Sep}(\mathcal{X}_1, \ldots, \mathcal{X}_m)^*$ contains the set of *entanglement witnesses* in the theory of entanglement, see [151].) Moreover, the use of "maximum" and "minimum" is justified in the above programs since $\overline{\rho} = \frac{I_{\mathcal{X}}}{\dim(\mathcal{X})}$ and $(\overline{t}, \overline{W}) = (2, 2I_{\mathcal{X}} - C)$ are strictly feasible solutions for (P) and (D), respectively [124, 125, 126] (i.e. strong duality (Theorem 4.8) holds).

Given two protocols, the corresponding cone programs are completely specified by Arthur's POVM corresponding to outcome *accept* and the underlying cone:

$$(C_1, \mathrm{Sep}\left(\mathcal{X}_1, \ldots, \mathcal{X}_m\right)) \text{ and } (C_2, \mathrm{Sep}\left(\mathcal{Y}_1, \ldots, \mathcal{Y}_m\right)), \tag{4.33}$$

while the parallel repetition protocol is specified by $(C_1 \otimes C_2, \mathrm{Sep}\left(\mathcal{X}_1 \otimes \mathcal{Y}_1, \ldots, \mathcal{X}_m \otimes \mathcal{Y}_m\right))$.

To show Theorem 4.3, note first that if ρ_1 and ρ_2 are optimal solutions of the primal problems associated with the two individual protocols, then $\rho_1 \otimes \rho_2$ is a feasible solution of the primal problem associated with the parallel repetition protocol. Therefore the success

probability of the parallel repetition is at least the product of the success probabilities of the individual protocols. We now show that *no* other strategy for the prover can perform better than this honest strategy. To do so, we demonstrate a feasible solution for the dual problem associated with the parallel repetition protocol attaining the same objective value.

More formally, let (t_1, W_1) and (t_2, W_2) be respective dual optimal solutions corresponding to two protocols. We show that $(t_1 \cdot t_2, W)$ is a dual feasible solution corresponding to the two-fold repetition of protocols for some choice of $W \in \mathrm{Sep}(\mathcal{X}_1 \otimes \mathcal{Y}_1, \ldots, \mathcal{X}_m \otimes \mathcal{Y}_m)^*$. To do so, we first require the following lemma.

Lemma 4.9. *For complex Euclidean spaces $\mathcal{X}_1, \ldots, \mathcal{X}_m, \mathcal{Y}_1, \ldots, \mathcal{Y}_m$:*

- $\mathrm{Sep}(\mathcal{X}_1, \ldots, \mathcal{X}_m)^* \otimes \mathrm{Sep}(\mathcal{Y}_1, \ldots, \mathcal{Y}_m) \subseteq \mathrm{Sep}(\mathcal{X}_1 \otimes \mathcal{Y}_1, \ldots, \mathcal{X}_m \otimes \mathcal{Y}_m)^*$*, and*
- $\mathrm{Sep}(\mathcal{X}_1, \ldots, \mathcal{X}_m) \otimes \mathrm{Sep}(\mathcal{Y}_1, \ldots, \mathcal{Y}_m)^* \subseteq \mathrm{Sep}(\mathcal{X}_1 \otimes \mathcal{Y}_1, \ldots, \mathcal{X}_m \otimes \mathcal{Y}_m)^*$.

Proof. We prove the first condition as the second is similar. Fix $W \in \mathrm{Sep}(\mathcal{X}_1, \ldots, \mathcal{X}_m)^*$ and $C \in \mathrm{Sep}(\mathcal{Y}_1, \ldots, \mathcal{Y}_m)$. Then for $S \in \mathrm{Sep}(\mathcal{X}_1 \otimes \mathcal{Y}_1, \ldots, \mathcal{X}_m \otimes \mathcal{Y}_m)$, we have

$$\langle W \otimes C, S \rangle = \langle W, \mathrm{Tr}_{\mathcal{Y}} \left[S(I_{\mathcal{X}} \otimes C) \right] \rangle \geq 0, \tag{4.34}$$

if $\mathrm{Tr}_{\mathcal{Y}} \left[S(I_{\mathcal{X}} \otimes C) \right] \in \mathrm{Sep}(\mathcal{X}_1, \ldots, \mathcal{X}_m)$. To this end, let

$$S = \sum_{i=1}^{k} \bigotimes_{l=1}^{m} \rho_i(l) \qquad \text{and} \qquad C = \sum_{j=1}^{k'} \bigotimes_{l=1}^{m} \sigma_j(l), \tag{4.35}$$

where $\rho_i(l) \in \mathrm{Pos}\left(\mathcal{X}_l \otimes \mathcal{Y}_l\right)$ and $\sigma_j(l) \in \mathrm{Pos}\left(\mathcal{Y}_l\right)$ for all $i \in [k]$, $j \in [k']$, and $l \in [m]$. Now we can write $\mathrm{Tr}_{\mathcal{Y}} \left[S \left(I_{\mathcal{X}} \otimes C \right) \right]$ as

$$\mathrm{Tr}_{\mathcal{Y}} \left[\left(\sum_{i=1}^{k} \bigotimes_{l=1}^{m} \rho_i(l) \right) \left(I_{\mathcal{X}} \otimes \sum_{j=1}^{k'} \bigotimes_{l=1}^{m} \sigma_j(l) \right) \right] \quad = \tag{4.36}$$

$$\sum_{i=1}^{k} \sum_{j=1}^{k'} \bigotimes_{l=1}^{m} \mathrm{Tr}_{\mathcal{Y}_l} \left[\rho_i(l) \left(I_{\mathcal{X}_l} \otimes \sigma_j(l) \right) \right].$$

Hence, $\mathrm{Tr}_{\mathcal{Y}} \left[S \left(I_{\mathcal{X}} \otimes C \right) \right] \in \mathrm{Sep}\left(\mathcal{X}_1, \ldots, \mathcal{X}_m \right)$ since $\mathrm{Tr}_{\mathcal{Y}_l} \left[\rho_i(l) \left(I_{\mathcal{X}_l} \otimes \sigma_j(l) \right) \right]$ is positive semidefinite for all i, j, l. The latter follows since for positive semidefinite $A_{\mathcal{X} \otimes \mathcal{Y}}$ and $B_{\mathcal{Y}}$,

$$\mathrm{Tr}_{\mathcal{Y}}(A_{\mathcal{X} \otimes \mathcal{Y}} I_{\mathcal{X}} \otimes B_{\mathcal{Y}}) = \mathrm{Tr}_{\mathcal{Y}}(I_{\mathcal{X}} \otimes B_{\mathcal{Y}}^{\frac{1}{2}} A_{\mathcal{X} \otimes \mathcal{Y}} I_{\mathcal{X}} \otimes B_{\mathcal{Y}}^{\frac{1}{2}}) \succeq 0, \tag{4.37}$$

which follows since $C^\dagger D C \succeq 0$ if $D \succeq 0$. This concludes the proof. □

We use Lemma 4.9 to construct two operators in $\mathrm{Sep}(\mathcal{X}_1 \otimes \mathcal{Y}_1, \ldots, \mathcal{X}_m \otimes \mathcal{Y}_m)^*$, the appropriate convex combination of which is the dual feasible solution we are seeking. Specifically, observe first that since for the two instances of the SepQMA(m) protocol, we have $C_1 \in \mathrm{Sep}\,(\mathcal{X}_1, \ldots, \mathcal{X}_m)$ and $C_2 \in \mathrm{Sep}\,(\mathcal{Y}_1, \ldots, \mathcal{Y}_m)$, and since $I_\mathcal{X}$ and $I_\mathcal{Y}$ are fully separable operators, it follows that

$$s_1 I_\mathcal{X} + C_1 \in \mathrm{Sep}(\mathcal{X}_1, \ldots, \mathcal{X}_m) \qquad \text{and} \qquad s_2 I_\mathcal{Y} + C_2 \in \mathrm{Sep}(\mathcal{Y}_1, \ldots, \mathcal{Y}_m) \tag{4.38}$$

for all $s_1, s_2 \geq 0$. Using Lemma 4.9, we thus obtain operators

$$(t_1 I_\mathcal{X} - C_1) \otimes (t_2 I_\mathcal{Y} + C_2) \in \mathrm{Sep}(\mathcal{X}_1 \otimes \mathcal{Y}_1, \ldots, \mathcal{X}_m \otimes \mathcal{Y}_m)^* \tag{4.39}$$

and

$$(t_1 I_\mathcal{X} + C_1) \otimes (t_2 I_\mathcal{Y} - C_2) \in \mathrm{Sep}(\mathcal{X}_1 \otimes \mathcal{Y}_1, \ldots, \mathcal{X}_m \otimes \mathcal{Y}_m)^*. \tag{4.40}$$

Here we have used the fact that $W = t_1 I_\mathcal{X} - C_1 \in \mathrm{Sep}(\mathcal{X}_1, \ldots, \mathcal{X}_m)^*$ since (t_1, W) is by assumption the optimal dual solution for the first protocol (and similarly for the second protocol). Since $\mathrm{Sep}(\mathcal{X}_1 \otimes \mathcal{Y}_1, \ldots, \mathcal{X}_m \otimes \mathcal{Y}_m)^*$ is a convex cone, it follows that the average of Equations (4.39) and (4.40) yields the desired operator

$$W := t_1 \cdot t_2\, I_{\mathcal{X}\otimes\mathcal{Y}} - C_1 \otimes C_2 \in \mathrm{Sep}(\mathcal{X}_1 \otimes \mathcal{Y}_1, \ldots, \mathcal{X}_m \otimes \mathcal{Y}_m)^*. \tag{4.41}$$

We conclude that $(t_1 \cdot t_2, W)$ is a feasible solution of the dual problem associated with parallel repetition of protocols with objective value $t_1 \cdot t_2$ as desired. This concludes the proof of Theorem 4.3.

Acknowledgements for this chapter. We thank Richard Cleve, Tsuyoshi Ito, Iordanis Kerenidis, Ashwin Nayak, Oded Regev, and Levent Tunçel for insightful discussions. We also thank LIAFA, Paris for their hospitality, where part of this work was completed.

Chapter 5

Signatures of non-classicality in mixed-state quantum computation

This chapter is based on [78]:

A. Datta and S. Gharibian. Signatures of nonclassicality in mixed-state quantum computation. *Physical Review A*, 79:042325, 2009, DOI: 10.1103/PhysRevA.79.042325,

In this chapter, we investigate signatures of non-classicality in quantum states, in particular, those involved in the DQC1 model of mixed-state quantum computation [174]. To do so, we consider two known non-classicality criteria. The first quantifies disturbance of a quantum state under locally noneffective unitary operations (LNU), which are local unitaries acting invariantly on a subsystem. The second quantifies measurement induced disturbance (MID) in the eigenbasis of the reduced density matrices. We study the role of both figures of non-classicality in the exponential speedup of the DQC1 model and compare them *vis-a-vis* the interpretation provided in terms of quantum discord. In particular, we prove that a non-zero quantum discord implies a non-zero shift under LNUs. We also use the MID measure to study the locking of classical correlations [87] using two mutually unbiased bases (MUB). We find the MID measure to exactly correspond to the number of locked bits of correlation.

5.1 Introduction and results

A thorough understanding of classical and quantum correlations underlies their successful exploitation in quantum information science. Characterizing the relative roles and abilities of these two forms of correlations in performing specific computational and information processing tasks would be a valuable advance in the field. Substantial progress in this direction has already been achieved. The role of entangled states in quantum information processing and computing is quite well studied. Jozsa and Linden [160] showed that multipartite entanglement must grow unboundedly with the problem size if a pure-state quantum computation is to attain an exponential speedup over its classical counterpart. In the context of information processing, Masanes has shown [192] that all bipartite entangled states can enhance the teleporting power of some other state. In spite of these successes, there are instances of quantum computations where the quantum advantage cannot be attributed to entanglement. Meyer has presented a quantum search algorithm that uses no entanglement [194]. Instances are also known of oracle based problems that can be solved without entanglement, yet with certain advantages over the best known classical algorithms [47, 166].

Given this scenario, it becomes a logical necessity to study the essentialness of entanglement in quantum information science. A realistic motivation is that provided by mixed-state quantum computation. Pure states in a quantum computation inevitably get mixed due to decoherence. One way to address this issue would be to study the prospects of quantum computational speedup with mixed states themselves [25]. NMR quantum computation provides a good scenario for this. As a simplified model for this, Knill and Laflamme proposed the DQC1 or the 'power of one qubit' model [174]. Though not believed to be as powerful as a pure-state quantum computer, it is believed to provide an exponential speedup over the best known classical algorithm for estimating the normalized trace of a unitary matrix. The DQC1 model was found to have a limited amount of (bipartite) entanglement that does not increase with the system size. Additionally, for certain parameter settings, there is no distillable entanglement present whatsoever, and yet the model retains its exponential advantage. In this latter case the state has a positive partial transpose, and thus possesses, at most, just bound entanglement [77]. Looking for a more satisfactory explanation for the exponential speedup, the quantum discord [203, 138] was calculated, of which the amount found was a constant fraction of the maximum possible [80], regardless of the parameter settings for the model. In this chapter, we study two alternative methods of studying the quantum behavior of quantum computational and information tasks.

Our results: This chapter studies the non-classical correlations found in the DQC1 states for trace estimation, as well as those used in the locking of classical correlations [87], with respect to two quantification schemes abbreviated as LNU and MID.

1. Locally noneffective unitaries (LNU). *Locally noneffective unitary* operations (LNU) have previously been studied with the aim of developing an entanglement detection criterion [102, 107] (see Section for a definition 5.2). Here, we study whether LNU can be used to quantify non-classicality, motivated by the disturbance of a quantum state under unitary operations. Specifically, we employ LNU in analyzing the DQC1 model, which has previously been studied using the quantum discord. Thus, we compare these two certificates of non-classicality, with the aim of contrasting *disturbance under measurement* with *disturbance under unitary operations*. We also study a mixed-state task in the setting of quantum communication known as *locking* [87], which uses two mutually unbiased bases (MUB) to *lock* classical correlations in a quantum state. For both tasks, we find that LNU do not indicate a high level of correlations.

2. Measurement-Induced Disturbance (MID). We then study the DQC1 model using the Measurement-Induced Disturbance (MID) measure [185] in Section 5.5. Regarding the MID measure, in Reference [185], a preliminary analysis of the DQC1 model was begun. Here, we extend this analysis to the entire parameter range for the DQC1 model, including those which limit the DQC1 state to being at most bound entangled. This latter case is of particular interest due to the lack of distillable entanglement. We also study the task of locking. For the latter, the value of the MID measure is exactly the number of locked bits of correlation in the state.

Discussion. With regards to the LNU distance, we find (Equation (5.11)) that there is little non-classicality in the $n+1$ qubit DQC1 state. This behavior is very similar to that of negativity [241] in the DQC1 model which was used to characterize its entanglement [77]. The crucial difference is that the bipartite split chosen in Section 5.3 is separable, and therefore exhibits no entanglement at all. As the LNU distance vanishes exponentially quickly with growing n, one is hard-pressed to relegate the role of the resource exponentially speeding up the DQC1 model to it. Similarly, the LNU distance suggests vanishing non-classicality in the case of locking of classical correlations in quantum states.

We find the MID measure, on the other hand, to be considerably more satisfactory. The zero-entanglement split in the DQC1 model is shown to have a non-zero amount of non-classicality as per the MID measure. The magnitude of this measure, as shown in Figure (5.4), is a constant fraction of its maximum possible value. Further, the MID

measure performs well in quantifying non-classicality in the scenario of locking classical correlations in quantum states. Further studies in this direction are required before a comprehensive conclusion can be reached.

Organization of chapter. We begin in Section 5.2 by defining LNU. Section 5.3 studies LNU in the DQC1 model. Section 5.4 shows a one-way relationship between LNU and the quantum discord. In Section 5.5, we define the MID measure, and use it to study the DQC1 model in Section 5.5.1. In Section 5.5.2, we study both MID and LNU in the context of locking.

Notation. Throughout this chapter, for $\mathcal{D}(\mathcal{A}\otimes\mathcal{B})$ we denote the dimensions of $\mathcal{A}$ and $\mathcal{B}$ as M and N, respectively. All designations of a density matrix without any subscripts refers to a bipartite state. For example, ρ stands for ρ_{AB}.

5.2 Locally noneffective unitary (LNU) operations

We begin by introducing locally noneffective unitary operations (LNU), first proposed under the name local *cyclic* operations [102]. For this, consider a bipartite quantum state $\rho\in\mathcal{D}(\mathcal{A}\otimes\mathcal{B})$, shared between A and B such that $\rho_A=\mathrm{Tr}_B(\rho)$ and $\rho_B=\mathrm{Tr}_A(\rho)$. Suppose now that Alice performs a local unitary U_A that does not change her subsystem, that is, $\rho_A=U_A\rho_A U_A^\dagger$, or equivalently

$$[\rho_A,U_A]=0. \tag{5.1}$$

This action can, however, affect the state of the total system, such that if we define $\rho_f:=(U_A\otimes I_B)\rho(U_A\otimes I_B)^\dagger$, it is possible that $\rho\neq\rho_f$. Unitaries satisfying Equation (5.1) are called LNU [102]. To quantify the difference between ρ and ρ_f, we use

$$\mathrm{d}_{\max}(\rho):=\max_{\substack{U_A:\\ [\rho_A,U_A]=0}}\frac{1}{\sqrt{2}}\|\,\rho-\rho_f\,\|_{\mathrm{F}}=\max_{\substack{U_A:\\ [\rho_A,U_A]=0}}\sqrt{\mathrm{Tr}(\rho^2)-\mathrm{Tr}(\rho\rho_f)}, \tag{5.2}$$

where $\|\,A\,\|_{\mathrm{F}}=\sqrt{\mathrm{Tr}(A^\dagger A)}$ denotes the Frobenius norm. From the latter expression, it is clear that $0\leq\mathrm{d}_{\max}(\rho)\leq 1$.

For any product state $\rho_{prod}:=\rho_A\otimes\rho_B$, $\mathrm{d}_{\max}(\rho_{prod})=0$. Closed form expressions for $\mathrm{d}_{\max}(\rho)$ are known for (pseudo)pure states and Werner states [107]. As with the quantum discord, it is possible to have $\mathrm{d}_{\max}(\rho_{sep})>0$ for certain separable states, implying $\mathrm{d}_{\max}(\rho)$

is not a non-locality measure. Recall that a separable state $\rho_{sep} \in \mathcal{D}(\mathcal{A} \otimes \mathcal{B})$ is defined as one of the form

$$\rho_{sep} := \sum_k p_k |a_k\rangle\langle a_k| \otimes |b_k\rangle\langle b_k|, \tag{5.3}$$

where $\sum_k p_k = 1$, and the $|a_k\rangle \in \mathcal{A}$ and $|b_k\rangle \in \mathcal{B}$ are vectors of Euclidean norm 1. For two-qubit separable states, the maximum LNU distance attainable is [102]

$$\mathrm{d}_{\max}(\rho_{sep}) \leq \frac{1}{\sqrt{2}}. \tag{5.4}$$

As an illustration, the maximum LNU distance for the two-qubit isotropic state,

$$\rho_{iso} = \frac{1-z}{4} I_4 + z|\Psi\rangle\langle\Psi|, \qquad z \in [0,1] \tag{5.5}$$

where $|\Psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, is given by $\mathrm{d}_{\max}(\rho_{iso}) = z$ [107]. By Equation (5.4), we can conclude that the two-qubit isotropic state is entangled for $z > 1/\sqrt{2}$. The partial transpose test, which in this case is necessary and sufficient, shows that this state is actually entangled for all $z > 1/3$, showing that the LNU distance is weaker at detecting entangled states than the former.

We remark that we have restricted our attention here to the case where the LNU is applied to subsystem A of ρ. Let us derive a simple upper bound on $\mathrm{d}_{\max}(\rho)$ which holds regardless of which target subsystem we choose, and which proves useful throughout this chapter.

Theorem 5.1. *For any $\rho \in \mathcal{D}(\mathcal{A} \otimes \mathcal{B})$,*

$$\mathrm{d}_{\max}(\rho) \leq \sqrt{2\left(\mathrm{Tr}(\rho^2) - \frac{1}{MN}\right)}. \tag{5.6}$$

Proof. Since $\left\| \rho - \frac{I}{MN} \right\|_{\mathrm{F}}$ is invariant under unitary operations, we have via the triangle inequality that:

$$\| \rho - \rho_f \|_{\mathrm{F}} \leq \left\| \rho - \frac{I}{MN} \right\|_{\mathrm{F}} + \left\| \frac{I}{MN} - \rho_f \right\|_{\mathrm{F}} = 2 \left\| \rho - \frac{I}{MN} \right\|_{\mathrm{F}} = 2\sqrt{\mathrm{Tr}(\rho^2) - \frac{1}{MN}}$$

Substituting this expression in Equation (5.2) gives the desired result. □

Thus, if the purity of a state ρ strictly decreases as a function of the dimension, then $\mathrm{d}_{\max}(\rho) \to 0$ as $MN \to \infty$.

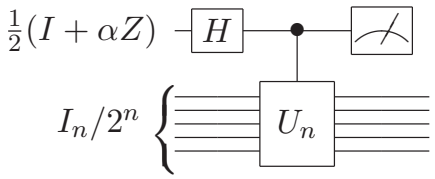


Figure 5.1: The DQC1 circuit

5.3 LNU in the DQC1 model

We now study the non-classical features of the DQC1 model of quantum computation, as quantified by $\mathrm{d}_{\max}(\rho)$. The $n+1$ qubit DQC1 state for given unitary $U_n \in \mathcal{U}(\mathcal{B}^{\otimes n})$, as demonstrated in Figure (5.1), is given by [77]

$$\rho_{DQC1} = \frac{1}{2^{n+1}} \begin{pmatrix} I_n & \alpha U_n^\dagger \\ \alpha U_n & I_n \end{pmatrix}. \tag{5.7}$$

We will consider the top qubit to be system A on which our local unitary acts and the remaining n qubits as system B. The reduced state is then

$$\rho_A = \mathrm{Tr}_B(\rho_{DQC1}) = \frac{1}{2} \begin{pmatrix} 1 & \alpha\tau^* \\ \alpha\tau & 1 \end{pmatrix} \tag{5.8}$$

with $\tau = \mathrm{Tr}(U_n)/2^n$. For an arbitrary SU(2) unitary U_A acting on A, which we characterize as

$$U_A = \begin{pmatrix} e^{i\phi}\cos\theta & e^{i\chi}\sin\theta \\ -e^{-i\chi}\sin\theta & e^{-i\phi}\cos\theta \end{pmatrix}, \tag{5.9}$$

the LNU condition of Equation (5.1) requires that $\chi = \frac{\pi}{2} - \arg(\tau)$ and either $\phi = 0$ or $\theta = \pi/2$. Both cases lead to the same final expression, so set $\phi = 0$. Via Equation (5.2) and simple algebra, we hence have

$$\mathrm{d}(\rho_{DQC1}, \theta) = \frac{\alpha \sin\theta}{2^{(n+1)/2}} \sqrt{1 - \frac{\mathrm{Re}(\mathrm{Tr}(e^{-2i\arg\tau} U_n^2))}{2^n}}. \tag{5.10}$$

The now trivial maximization over all θ gives

$$\mathrm{d}_{\max}(\rho_{DQC1}) = \frac{\alpha}{2^{(n+1)/2}} \sqrt{1 - \frac{\mathrm{Re}(\mathrm{Tr}(e^{-2i\arg\tau} U_n^2))}{2^n}} \leq \frac{\alpha}{2^{n/2}}. \tag{5.11}$$

Here, we have used the rough estimate $\mathrm{Re}(\mathrm{Tr}(e^{2i\arg\tau}U_n^2)) \geq -2^n$. For a two-qubit pure state ($n=1, \alpha=1$), we thus have $\mathrm{d}_{\max}(\rho_{DQC1}) \leq 1/\sqrt{2}$, which conforms with Equation (5.4). A typical instance of the DQC1 circuit is provided by that of a random unitary U_n in the DQC1 circuit of Figure (5.1). For such instances of large enough Haar distributed unitaries, $\mathrm{Tr}(U_n^2)$ is bounded above by a constant with high probability [85]. Thus, the second term inside the square root in Equation (5.11) is approximately zero, and

$$\mathrm{d}_{\max}(\rho_{DQC1}) \approx \frac{\alpha}{2^{(n+1)/2}}. \tag{5.12}$$

This shows that the DQC1 state experiences very little disturbance under LNU, and in fact this disturbance vanishes asymptotically as n grows. As discussed in the introduction, it would appear that the quantum discord is better suited [80] to quantifying non-classicality in the DQC1 model. This, however, raises the question of how the discord and LNU distance are related, and whether the paradigms of 'disturbance under measurement' and 'disturbance under unitary operations' lead to differing notions of non-classicality. We explore these questions in the following section.

Before closing, for completeness, we invoke Theorem (5.1) to show that the LNU distance is exponentially decreasing for *any* other choice of bi-partitions A and B of the qubits in ρ_{DQC1}. In fact, since

$$\mathrm{Tr}(\rho_{DQC1}^2) = \frac{1+\alpha^2}{2^{n+1}}, \tag{5.13}$$

Theorem (5.1) immediately gives the same upper bound of Equation (5.11).

5.4 Quantum discord *vs* LNU distance

Motivated by the fact that both the quantum discord and the LNU distance are aimed at capturing the non-classical features in a quantum state via an induced disturbance, we seek an answer to the question of whether one implies the other in any sense or not. Here, we show that non-zero quantum discord implies a non-zero LNU distance, but that the converse is not necessarily true. We begin by recalling the definition of quantum discord.

Given a quantum state $\rho \in \mathcal{D}(\mathcal{A}\otimes\mathcal{B})$, recall from Section 1.6.2 that the quantum discord [203] is defined as

$$\delta(\rho) := S(\rho_A) - S(\rho) + \min_{\{\Pi_j^A\}} S\left(\rho_{B|\{\Pi_j^A\}}\right) \tag{5.14}$$

for $\left\{\Pi_j^A\right\}$ a rank-one projective measurement, and for

$$S\left(\rho_{B|\left\{\Pi_j^A\right\}}\right) := \sum_j p_j S\left((\Pi_j^A \otimes I^B)\rho(\Pi_j^A \otimes I^B)\Big/ p_j\right), \tag{5.15}$$

where $p_j = \mathrm{Tr}(\Pi_j^A \otimes I^B \rho)$. Intuitively, quantum discord captures purely quantum correlations in a quantum state. This is distinct from entanglement in the case of mixed states. For pure states, quantum discord reduces to the von Neumann entropy of the reduced density matrix, which is a measure of entanglement. On the other hand, it is possible for mixed separable states to have non-zero quantum discord. The main theorem concerning the discord that we require here is the following.

Theorem 5.2 (Ollivier and Zurek [203])**.** *For $\rho \in \mathcal{D}(\mathcal{A} \otimes \mathcal{B})$, $\delta(\rho) = 0$ if and only if $\rho = \sum_j (\Pi_j^A \otimes I^B)\rho(\Pi_j^A \otimes I^B)$, for some complete set of rank one projectors $\left\{\Pi_j^A\right\}$.*

We now show the following.

Theorem 5.3. *For $\rho \in \mathcal{D}(\mathcal{A} \otimes \mathcal{B})$, if $\delta(\rho) > 0$, then $\mathrm{d}_{\max}(\rho) > 0$.*

Proof. We begin by writing ρ in Fano form [91], i.e.

$$\rho = \frac{1}{MN}(I^A \otimes I^B + \boldsymbol{r}^A \cdot \boldsymbol{\sigma}^A \otimes I^B + I^A \otimes \boldsymbol{r}^B \cdot \boldsymbol{\sigma}^B + \sum_{s=1}^{M^2-1}\sum_{t=1}^{N^2-1} T_{st}\sigma_s^A \otimes \sigma_t^B). \tag{5.16}$$

Here, $\boldsymbol{\sigma}^A$ denotes a (M^2-1)-component vector of traceless orthogonal Hermitian basis elements (which generalize the Pauli spin operators), $\boldsymbol{r}^A$ is the (M^2-1)-dimensional Bloch vector for subsystem A with $r_s^A = \frac{M}{2}\mathrm{Tr}(\rho_A \sigma_s^A)$, and T is a real matrix known as the correlation matrix with entries $T_{st} = \frac{MN}{4}\mathrm{Tr}(\sigma_s^A \otimes \sigma_t^B \rho)$. The definitions for subsystem B are analogous.

An explicit construction for the basis elements σ_i for $M \geq 2$ is given as follows [139]. Define $\{\sigma_i\}_{i=1}^{M^2-1} = \{U_{pq}, V_{pq}, W_r\}$, such that for $1 \leq p < q \leq M$ and $1 \leq r \leq M-1$, and $\{|k\rangle\}_{k=1}^M$ some complete orthonormal basis for $\mathcal{A}$:

$$U_{pq} = |p\rangle\langle q| + |q\rangle\langle p| \tag{5.17}$$

$$V_{pq} = -i|p\rangle\langle q| + i|q\rangle\langle p| \tag{5.18}$$

$$W_r = \sqrt{\frac{2}{r(r+1)}}\left(\sum_{k=1}^{r}|k\rangle\langle k| - r|r+1\rangle\langle r+1|\right). \tag{5.19}$$

In our ensuing discussion, without loss of generality, we fix the choice of basis $\{|k\rangle\}_{k=1}^{M}$ above as the eigenbasis of ρ_A. (Note that the set of orthonormal eigenvectors of ρ_A will not be unique if the eigenvalues of ρ_A are degenerate. Hence, we fix some choice of eigenbasis for ρ_A as the "canonical" choice to be referred to throughout the rest of our discussion.)

Assume now that $\delta(\rho) > 0$. Then, any choice of complete measurement $\{\Pi_j^A\}$ must disturb ρ, i.e. by Theorem 5.2, if we define

$$\rho_f := \sum_{j=1}^{M} (\Pi_j^A \otimes I)\rho(\Pi_j^A \otimes I), \tag{5.20}$$

then $\rho_f \neq \rho$. Henceforth, when we discuss the action of $\{\Pi_j^A\}$ on ρ_A, we are referring to the state $\sum_{j=1}^{M} \Pi_j^A \rho_A \Pi_j^A$. Now, let $\{\Pi_j^A\}$ be a complete projective measurement onto the eigenbasis of ρ_A. Then, $\{\Pi_j^A\}$ acts invariantly on ρ_A, and thus must alter the last term in Equation (5.16) to ensure $\rho_f \neq \rho$. To see this, recall that one can write $\rho_A = \frac{1}{M}(I^A + \boldsymbol{r}^A \cdot \boldsymbol{\sigma}^A)$, from which it follows that if $\{\Pi_j^A\}$ acts invariantly on ρ_A, then it also acts invariantly on $\boldsymbol{r}^A \cdot \boldsymbol{\sigma}^A$ from Equation (5.16). Since all basis elements $\sigma_s^A \in \{W_r\}_r$ are diagonal, it follows that there must exist some $T_{st} \neq 0$ such that $\sigma_i^A \in \{U_{pq}, V_{pq}\}_{pq}$. We now use this fact to construct a LNU U^A achieving $\mathrm{d}(\rho, U_A) > 0$.

Define unitary U^A as diagonal in the eigenbasis of ρ_A, i.e. $U^A = \sum_{k=1}^{M} e^{i\theta_k}|k\rangle\langle k|$, with eigenvalues to be chosen as needed. Then, $[U^A, \rho_A] = 0$ by construction, and so $U^A \otimes I^B$ must alter T through its action on ρ to ensure $\rho_f \neq \rho$. Focusing on the last term from Equation (5.16), we thus have:

$$\sum_{s=1}^{M^2-1}\sum_{t=1}^{N^2-1} T_{st} U^A \sigma_s^A U^{A\dagger} \otimes \sigma_t^B = \sum_{s=1}^{M^2-1}\sum_{t=1}^{N^2-1} T_{st} \left(\sum_{m=1}^{M}\sum_{n=1}^{M} e^{i(\theta_m - \theta_n)} \langle m|\sigma_s^A|n\rangle |m\rangle\langle n| \right) \otimes \sigma_t^B$$

Analyzing each σ_s^A case by case, we find, for some $1 \leq p < q \leq M$ or $1 \leq r \leq M-1$:

$$\sum_{m=1}^{M}\sum_{n=1}^{M} e^{i(\theta_m - \theta_n)} \langle m|\sigma_s|n\rangle |m\rangle\langle n| = \begin{cases} \cos(\theta_p - \theta_q)U_{pq} - \sin(\theta_p - \theta_q)V_{pq} & \text{if } \sigma_s = U_{pq} \\ \sin(\theta_p - \theta_q)U_{pq} + \cos(\theta_p - \theta_q)V_{pq} & \text{if } \sigma_s = V_{pq} \\ W_r & \text{if } \sigma_s = W_r. \end{cases} \tag{5.21}$$

Denoting by T^f the T matrix for ρ_f, we have:

$$T^f_{st} = \begin{cases} \cos(\theta_p - \theta_q)T_{st} + \sin(\theta_p - \theta_q)T_{wt} \\ \qquad\qquad \text{if } \sigma_s = U_{pq}, \text{ where } \sigma_w = V_{pq} \\ \cos(\theta_p - \theta_q)T_{st} - \sin(\theta_p - \theta_q)T_{wt} \\ \qquad\qquad \text{if } \sigma_s = V_{pq}, \text{ where } \sigma_w = U_{pq} \\ T_{st} \qquad\qquad \text{if } \sigma_s = W_r. \end{cases} \tag{5.22}$$

Thus, if there exists an s such that $T_{st} \neq 0$ and $\sigma^A_s \in \{U_{pq}, V_{pq}\}_{pq}$, it follows that one can easily choose appropriate eigenvalues $e^{i\theta_p}$ and $e^{i\theta_q}$ for U^A such that $T^f \neq T$, implying $\mathrm{d}_{\max}(\rho) > 0$. By our argument above for $\delta(\rho) > 0$, such an s does in fact exist. □

To show that the converse of Theorem 5.3 does not hold, we present an example of a zero discord state that has non-zero LNU measure. Consider the two qubit separable state

$$\rho = \frac{1}{2}\left(\frac{I_2 + \boldsymbol{a}.\boldsymbol{\sigma}}{2} \otimes \frac{I_2 + \boldsymbol{b}.\boldsymbol{\sigma}}{2} + \frac{I_2 - \boldsymbol{a}.\boldsymbol{\sigma}}{2} \otimes \frac{I_2 - \boldsymbol{b}.\boldsymbol{\sigma}}{2}\right), \tag{5.23}$$

where $\| \boldsymbol{a} \|_2 = \| \boldsymbol{b} \|_2 = 1$. This state, by construction, has zero discord for a single qubit measurement on either A or B. To see this, consider the projective measurements

$$\left\{\frac{I_2 \pm \boldsymbol{a}.\boldsymbol{\sigma}}{2}\right\} \tag{5.24}$$

on A. Let us now study the LNU distance for this state, with the local unitary being applied to say A. Notice that $\rho_A = \rho_B = I_2/2$, and $\mathrm{Tr}(\rho^2) = 1/2$. The former implies that any local unitary on $\mathcal{A}$ can be chosen, as characterized by Equation (5.9). Let us for convenience parameterize $\boldsymbol{a} = (0, 0, 1)$ and $\boldsymbol{b} = (\sin\gamma\cos\delta, \sin\gamma\sin\delta, \cos\gamma)$. Then, some algebra leads to

$$\mathrm{Tr}(\rho\rho_f) = \frac{1}{2}\cos^2\theta, \tag{5.25}$$

whose minimum is 0, whereby

$$\mathrm{d}_{\max}(\rho) = \frac{1}{\sqrt{2}}. \tag{5.26}$$

We thus have an example of a class of separable, zero discord states which demonstrates a non-zero shift under LNU. In fact, it attains the maximum shift possible for two-qubit separable states. Hence, if one wishes to define notions of non-classicality in quantum states in terms of 'disturbance under measurement' versus 'disturbance under unitary operations',

and one chooses discord and the LNU distance as canonical quantifiers of such effects, respectively, then the resulting respective notions of non-classicality are not equivalent. As we have shown in Theorem 5.3, however, the quantum discord is a stronger notion of non-classicality than the LNU criterion.

5.5 Measuring correlations via measurement-induced disturbance

The measure we intend to use in this section was presented by Luo in [185]. It relies on the disturbance of a quantum system under a generic measurement. In that sense, it is similar in spirit to quantum discord, but not quite. In the case of quantum discord, as per Equation (5.14), one maximizes over one-dimensional projective measurements on one of the subsystems. For the measure used here, which we will call the Measurement-Induced Disturbance (MID) measure, one performs measurements on *both* the subsystems, with the measurements being given by projectors onto the eigenvectors of the reduced subsystems. Then the MID measure of quantum correlations for a quantum state $\rho \in \mathcal{D}(\mathcal{A}\otimes\mathcal{B})$ is given by [185]

$$\mathcal{M}(\rho) := \mathcal{I}(\rho) - \mathcal{I}(\mathcal{P}(\rho)) \tag{5.27}$$

where

$$\mathcal{P}(\rho) := \sum_{i=1}^{M}\sum_{j=1}^{N}(\Pi_i^A \otimes \Pi_j^B)\rho(\Pi_i^A \otimes \Pi_j^B). \tag{5.28}$$

Here $\{\Pi_i^A\}, \{\Pi_j^B\}$ denote rank one projections onto the eigenbases of ρ_A and ρ_B, respectively, and $\mathcal{I}(\sigma)$ is the quantum mutual information. The measurement induced by the local eigenvectors leaves the entropy of the reduced states invariant and is, in a certain sense, the least disturbing. Actually, this choice of measurement even leaves the reduced states invariant [185]. Interestingly, for pure states, both the quantum discord and the MID measure reduce to the von Neumann entropy of the reduced density matrix, which is a measure of bipartite entanglement. An advantage of the MID measure is that since no optimizations are involved, it is much easier to calculate in practice than the quantum discord or the LNU distance, which involve optimizations over projective measurements and local unitaries respectively. The corresponding disadvantage is that if the spectrum of either ρ_A or ρ_B is degenerate, there exist examples [259] where the MID measure is not necessarily well-defined, as the choice of local eigenbases is no longer unique. In this case, the value of the MID measure should be interpreted moreso as a rough estimate or *upper*

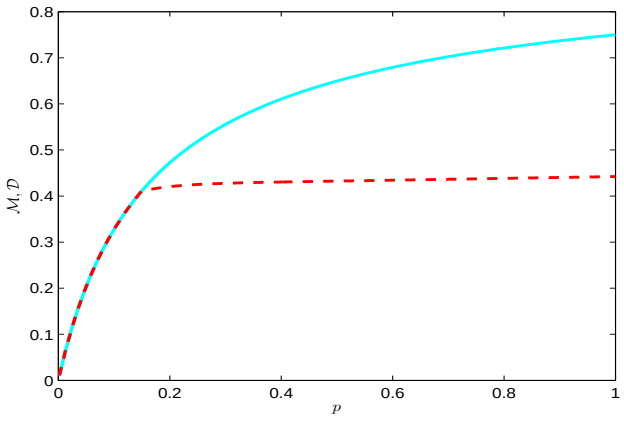


Figure 5.2: (Color online) The solid line is the MID measure $\mathcal{M}$ for the 2×4 Horodecki state from [147]. The dashed line is the quantum discord $\mathcal{D}$ for the same state [79]. The kink in the latter curve occurs at $p = 1/7$. We see here, as in the case of the DQC1 state, that the MID measure is greater than or equal to the quantum discord.

bound on the non-classicality of a state. We remark that for this reason, it may be more reasonable to consider a quantity

$$\mathcal{M}^*(\rho) := \mathcal{I}(\rho) - \max_{\{\Pi_i^A\},\{\Pi_j^B\}} \mathcal{I}(\mathcal{P}(\rho)), \tag{5.29}$$

where $\{\Pi_i^A\}, \{\Pi_j^B\}$ are again projections onto eigenbases of ρ_A and ρ_B, respectively. (A quantity similar to $\mathcal{M}^*(\rho)$ was considered in [259], except the maximization there is over *all* local POVMs. Also, note that it follows directly from the definition of $\mathcal{M}^*(\rho)$ that it is an upper bound on the *distillable entanglement potential* of ρ introduced in Chapter 7 (Equation (7.16)).) Computing $\mathcal{M}^*(\rho)$ is naturally much more difficult; we discuss $\mathcal{M}^*(\rho)$ in this section where appropriate in addition to our discussion of $\mathcal{M}(\rho)$.

To demonstrate the MID measure on a non-trivial example, we first consider the well-known Horodecki bound entangled state in $2 \otimes 4$ dimensions [147]. It is bound entangled

for all values of $0 \leq p \leq 1$, and the state is given as

$$\rho_H = \frac{1}{1+7p}\begin{pmatrix} p & 0 & 0 & 0 & 0 & p & 0 & 0 \\ 0 & p & 0 & 0 & 0 & 0 & p & 0 \\ 0 & 0 & p & 0 & 0 & 0 & 0 & p \\ 0 & 0 & 0 & p & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1+p}{2} & 0 & 0 & \frac{\sqrt{1-p^2}}{2} \\ p & 0 & 0 & 0 & 0 & p & 0 & 0 \\ 0 & p & 0 & 0 & 0 & 0 & p & 0 \\ 0 & 0 & p & 0 & \frac{\sqrt{1-p^2}}{2} & 0 & 0 & \frac{1+p}{2} \end{pmatrix}. \tag{5.30}$$

The projectors onto the eigenvectors of the reduced density matrices can be chosen as

$$\{\Pi_1^A, \Pi_2^A\} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}, \quad \text{and} \tag{5.31}$$

$$\{\Pi_1^B, \cdots, \Pi_4^B\} = \left\{ |\Psi^+\rangle\langle\Psi^+|, |\Psi^-\rangle\langle\Psi^-|, |\Phi^+\rangle\langle\Phi^+|, |\Phi^-\rangle\langle\Phi^-| \right\}. \tag{5.32}$$

where $|\Psi^\pm\rangle = (|1\rangle \pm |2\rangle)/\sqrt{2}$ and $|\Phi^\pm\rangle = (|0\rangle \pm |3\rangle)/\sqrt{2}$, with $\{|0\rangle, |1\rangle, |2\rangle, |3\rangle\}$ forming the computational basis for the second subsystem. Using these in Equation (5.28), we have

$$\mathcal{P}(\rho_H) = \frac{1}{1+7p}\begin{pmatrix} p & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1+p}{2} & 0 & 0 & \frac{\sqrt{1-p^2}}{2} \\ 0 & 0 & 0 & 0 & 0 & p & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p & 0 \\ 0 & 0 & 0 & 0 & \frac{\sqrt{1-p^2}}{2} & 0 & 0 & \frac{1+p}{2} \end{pmatrix}. \tag{5.33}$$

Note that this density matrix differs from ρ_H in that some of the off-diagonal terms p have vanished. We have computed the MID measure for ρ_H as $\mathcal{M}(\rho_H) = S(\mathcal{P}(\rho_H)) - S(\rho_H)$ and plotted it in Figure (5.2). In the same figure, we also plot the quantum discord for this state, when a measurement is made on the two-dimensional subsystem [79]. As we see, there are non-classical correlations in this state that are not distillable into maximally entangled Bell pairs.

As a comparison, we remark that for ρ_H, $\mathcal{M}^*(\rho_H)$ behaves similarly to $\mathcal{M}(\rho_H)$. To see this, note that only ρ_B has a degenerate eigenvalue, and this is on the space spanned by

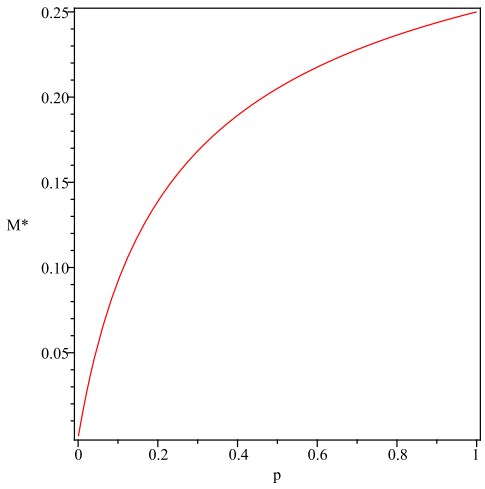


Figure 5.3: A plot of $\mathcal{M}^*$ for the 2×4 Horodecki state from [147].

Π_3^B and Π_4^B. Thus, in the minimization over local bases, one can more generally choose Π_3^B and Π_4^B to project onto an arbitrary basis for this space, $a|0\rangle + e^{i\theta} b|3\rangle$ and $b|0\rangle - e^{i\theta} a|3\rangle$ for $a, b, \theta \in \mathbb{R}$, respectively. The eigenvalues of $\mathcal{P}(\rho_H)$ are then (up to normalization)

$$\frac{1}{2}\left(p + 1 \pm \sqrt{1-p^2}\right), p, p, p\left(1 \pm |a|\,|b|\,\sqrt{2(1-\cos(2\theta))}\right), p\left(1 \pm |a|\,|b|\,\sqrt{2(1-\cos(2\theta))}\right). \tag{5.34}$$

In the expression $\mathcal{M}^*(\rho_H) = \min_{\{\Pi_i^A\},\{\Pi_j^B\}} S(\mathcal{P}(\rho_H)) - S(\rho_H)$, the entropy $S(\mathcal{P}(\rho_H))$ is thus minimized by choosing $a = b = 1/\sqrt{2}$ and $\theta = \pi/2$. A plot of the resulting value of $\mathcal{M}^*(\rho_H)$ is given in Figure 5.3.

5.5.1 MID measure in the DQC1 model

We now move on to calculate the MID measure in the DQC1 model. Our analysis extends that of [185], where only the case of $\alpha = 1$ was considered. Considering $\alpha < 1/2$ here will be of particular interest, due to the lack of distillable entanglement in the DQC1 state (in this regime, any bipartite split has a positive partial transpose). Consequently, we start with the $(n+1)$-qubit DQC1 state, given by Equation (5.7), wherefrom

$$\rho_A = \frac{1}{2}\begin{pmatrix} 1 & \alpha\tau^* \\ \alpha\tau & 1 \end{pmatrix} \quad \text{and} \quad \rho_B = I_n/2^n, \tag{5.35}$$

where recall $\tau = \mathrm{Tr}(U_n)/2^n$. The projectors onto ρ_A's eigenvectors can be chosen as

$$\{\Pi_0^A, \Pi_1^A\} = \{|\phi_0\rangle\langle\phi_0|, |\phi_1\rangle\langle\phi_1|\} \tag{5.36}$$

for $|\phi_0\rangle := (|0\rangle + e^{i\phi}|1\rangle)/\sqrt{2}$ and $|\phi_1\rangle := (|0\rangle - e^{i\phi}|1\rangle)/\sqrt{2}$, respectively, where $\tau = re^{i\phi}$ for $r = |\tau|$. Similarly, set $\{\Pi_j^B\} = \{|j\rangle\langle j|\}$ for $\{|j\rangle\}_{j=1}^{2^n}$ the computational basis. Using this, we can calculate

$$\begin{aligned}
\mathcal{P}(\rho_{DQC1}) &= \sum_{j=1}^{2^n}\sum_{k=0}^{1}(\Pi_k^A \otimes \Pi_j^B)\rho_{DQC1}(\Pi_k^A \otimes \Pi_j^B) && (5.37)\\
&= \frac{1}{2^{n+1}}\sum_{j=1}^{2^n}\sum_{k=0}^{1}\Pi_k^A \begin{pmatrix} 1 & \alpha\langle j|U_n^\dagger|j\rangle \\ \alpha\langle j|U_n|j\rangle & 1 \end{pmatrix}\Pi_k^A \otimes |j\rangle\langle j|. && (5.38)
\end{aligned}$$

Observing that

$$\Pi_k^A \begin{pmatrix} 1 & \alpha\langle j|U_n^\dagger|j\rangle \\ \alpha\langle j|U_n|j\rangle & 1 \end{pmatrix}\Pi_k^A = \left(1 + (-1)^k\alpha\,\mathrm{Re}(\langle j|U_n|j\rangle e^{-i\phi})\right)|\phi_k\rangle\langle\phi_k|, \tag{5.39}$$

we conclude that the spectrum of $\mathcal{P}(\rho_{DQC1})$ is given by

$$\boldsymbol{\lambda}[\mathcal{P}(\rho_{DQC1})] = \left\{\frac{1 \pm \Delta_j}{2^{n+1}}\right\} \tag{5.40}$$

for

$$\Delta_j := \alpha\,\mathrm{Re}\left(\langle j|U_n|j\rangle e^{-i\phi}\right) \tag{5.41}$$

and $j \in [2^n]$. Letting λ_k denote the kth entry of $\boldsymbol{\lambda}[\mathcal{P}(\rho_{DQC1})]$, the von Neumann entropy of this state is

$$\begin{aligned}
S(\mathcal{P}(\rho_{DQC1})) &= -\sum_{k=1}^{2^{n+1}}\lambda_k\log(\lambda_k) && (5.42)\\
&= n + 1 - \frac{1}{2^{n+1}}\sum_{j=1}^{2^n}\left(\log(1-\Delta_j^2) + \Delta_j\log\left(\frac{1+\Delta_j}{1-\Delta_j}\right)\right). && (5.43)
\end{aligned}$$

Now,

$$S(\rho_{DQC1}) = n + H_2\left(\frac{1-\alpha}{2}\right), \tag{5.44}$$

and since the entropies of the partial density matrices are invariant under the local measurements, we have

$$\begin{aligned}\mathcal{M}_{DQC1} &= \mathcal{I}(\rho_{DQC1}) - \mathcal{I}(\mathcal{P}(\rho_{DQC1})) && (5.45)\\ &= S(\mathcal{P}(\rho_{DQC1})) - S(\rho_{DQC1}) && (5.46)\\ &= 1 - H_2\left(\frac{1-\alpha}{2}\right) - \frac{1}{2^{n+1}}\sum_{j=1}^{2^n}\left(\log(1-\Delta_j^2) + \Delta_j \log\left(\frac{1+\Delta_j}{1-\Delta_j}\right)\right).\end{aligned}$$

For any unitary U_n, which is known in any implementation of the DQC1 circuit, the above quantity can be computed easily. Bounding this quantity more generally, however, is difficult. If, however, in the asymptotic limit of large n, $|\Delta_j| \to 0$ (as might intuitively be expected when U_n is a Haar distributed random unitary matrix, since then we might expect $|u_{jj}| \sim 1/2^{n/2}$), then the whole quantity within the summation in Equation (5.45) goes to zero. In this case,

$$\mathcal{M}_{DQC1} \sim 1 - H_2\left(\frac{1-\alpha}{2}\right). \tag{5.47}$$

One fact immediately notable is that the above expression for the MID measure is independent of n, for large n. The result for a $n = 5$ qubit Haar distributed random unitary matrix is shown in Figure (5.4). As is evident, despite the approximations used in the derivation of Equation (5.47) the asymptotic analytic expression matches the numerical result at $n = 5$ quite well. We remark that even for a version of $\mathcal{M}_{DQC1}$ where one minimizes over all local POVMs, the behavior one finds is quantitatively analogous to that of $\mathcal{M}_{DQC1}$ plotted in Figure 5.4 [259].

The MID measure for the DQC1 state across the bipartite split separating the top qubit from the rest is non-zero for all non-zero values of α. Across this split, the DQC1 state is strictly separable [77] and possesses no entanglement. Hence, one might propose the MID measure as a quantifier of the resource behind the quantum advantage in the DQC1 model [185]. Note that, as can be seen from Figure (5.4), the behavior of the MID measure is qualitatively quite similar to that of the quantum discord.

5.5.2 Non-classical correlations in quantum communication

We now use the MID measure to study the locking of classical correlations in quantum states. It has been shown [87] that there exist bipartite quantum states which contain a large amount of locked classical correlation which can be unlocked by a small amount of classical communication. More precisely, there exist $(2n + 1)$-qubit states for which the

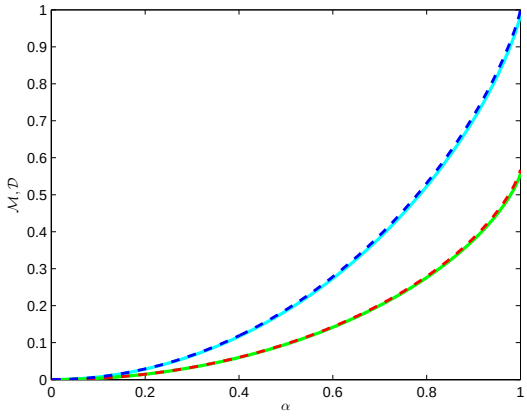


Figure 5.4: (Color online) The upper solid (cyan) line is the MID measure $\mathcal{M}$ (Equation (5.45)) for the DQC1 circuit for a $n = 5$ qubit Haar distributed random unitary matrix. The upper dashed (blue) line is the analytic expression for the MID measure for certain DQC1 states from Equation (5.47). The lower dashed (red) line shows the discord $\mathcal{D}$ in the DQC1 circuit with the same unitary. The lower solid (green) line shows the analytical expression of the quantum discord from [80]. All quantities are shown as functions of the purity α of the control qubit.

optimal classical mutual information between measurement results on the subsystems can be increased from $n/2$ bits to n bits via a single bit of classical communication. Despite the impossibility of this feat classically, the states used in the protocol are not entangled.

Here we use the MID measure to study this purely quantum phenomenon. To do so, we evaluate the former on a generalization of the state used in [87],

$$\rho = \frac{1}{md} \sum_{k=1}^{d} \sum_{t=1}^{m} (|k\rangle\langle k| \otimes |t\rangle\langle t|)_A \otimes (|b_k^t\rangle\langle b_k^t|)_B, \tag{5.48}$$

where the set of m orthonormal bases $\left\{ \{|b_k^t\rangle\}_{k=1}^{d} \right\}_{t=1}^{m}$ is mutually unbiased (MUB), i.e. $\forall_{t \neq t', i,j} \left| \langle b_i^t | b_j^{t'} \rangle \right| = 1/\sqrt{d}$. As in Reference [87], when $d = 2^n$ and $m = 2$, the initial correlations in this state amount to $n/2$ bits, and by Alice's sending one bit (the bit t) to Bob, they end up with $n + 1$ correlated bits. The state being separable, it has no entanglement. Consequently, we cannot ascribe to entanglement the advantage exhibited by this protocol.

To calculate the MID measure of this state, we need the reduced states given by

$$\rho_A = \frac{I_{md}}{md}, \quad \rho_B = \frac{I_d}{d}. \tag{5.49}$$

Choosing the local eigenvectors as the respective computational bases, we have that $\mathcal{P}(\rho)$ is simply the diagonal of ρ. Thus,

$$\boldsymbol{\lambda}[\mathcal{P}(\rho)] = \frac{1}{md}\{\underbrace{1, \cdots, 1}_{d}, \underbrace{1/d, \cdots, 1/d}_{(m-1)d^2}, \underbrace{0, 0, \cdots, 0}_{d(d-1)}\} \tag{5.50}$$

whereby

$$S(\mathcal{P}(\rho)) = \log m + \left(2 - \frac{1}{m}\right) \log d. \tag{5.51}$$

The spectrum of ρ is given by

$$\boldsymbol{\lambda}[\rho] = \frac{1}{md}\{\underbrace{1, 1, \cdots, 1}_{md}, \underbrace{0, 0, \cdots, 0}_{md(d-1)}\} \tag{5.52}$$

which leads to

$$S(\rho) = \log m + \log d. \tag{5.53}$$

Finally, we have

$$\mathcal{M}(\rho) = S(\mathcal{P}(\rho)) - S(\rho) = \left(1 - \frac{1}{m}\right) \log d, \tag{5.54}$$

which for $d = 2^n$ and $m = 2$ is the exactly equal to the gain attained by this scheme. Moreover, once Bob receives Alice's bit, the MID measure for their post-communication state drops to 0, the latter being diagonal in a local product basis. This suggests the possibility that the MID measure quantifies the non-classical (yet not entanglement-based) correlations in ρ which were initially locked. Moreover, we remark that for $d = 2^n$ and $m = 2$, we have $\mathcal{M}(\rho) = \mathcal{M}^*(\rho)$ — this follows directly from the result [87] that the mutual information of any classical distribution induced via local measurements on ρ is at most $(\log d)/2$.

A few remarks are in order. Equation (5.54) might suggest that a better locking effect may be possible for $m > 2$. However, explicit constructions to date using more than two MUBs have been unable to achieve superior locking [34], suggesting that the choice of construction for the MUBs plays an important role. In contrast, Equation (5.54) holds irrespective of the specific choice of MUBs. It is also known that if the bases above are

constructed using a large set of random unitaries chosen according to the Haar measure, then the classical mutual information in ρ between Alice and Bob can be brought down to a constant [137]. There is also numerical evidence (Appendix of Reference [87]) that the dimension of the systems may play a role in achieving better locking. Connections between locking and non-classical correlations have since been discovered in References [259, 49].

Finally, for completeness, we remark that $\mathrm{Tr}(\rho^2) = 1/(md)$, and so by Theorem 5.1, the LNU distance for ρ is bounded by

$$\mathrm{d}_{\max}(\rho) \leq \sqrt{\frac{2}{md}\left(1 - \frac{1}{d}\right)} \leq \sqrt{\frac{2}{md}}. \tag{5.55}$$

Thus, in contrast to the MID measure, the LNU distance once again reveals vanishing non-classicality with growing m or d.

Acknowledgements for this chapter. We thank Carl Caves and Anil Shaji for numerous stimulating discussions, as well as an anonymous referee for raising certain points that led to improvements in the paper this chapter is based on.

Chapter 6

Quantifying non-classicality with local unitary operations

This chapter is based on [106]:

S. Gharibian. Quantifying non-classicality with local unitary operations. Available at arXiv.org e-Print quant-ph/1202.1598v1, 2012.

In this chapter, we propose a measure of non-classical correlations in bipartite quantum states based on local unitary operations. We prove the measure is non-zero if and only if the quantum discord is non-zero; this is achieved via a new characterization of zero discord states in terms of the state's correlation matrix. Moreover, our scheme can be extended to ensure the same relationship holds even with a generalized version of quantum discord in which higher-rank projective measurements are allowed. We next derive a closed form expression for our scheme in the cases of Werner states and $(2 \times N)$-dimensional systems. The latter reveals that for $(2 \times N)$-dimensional states, our measure reduces to the geometric discord [75]. A connection to the CHSH inequality is shown. We close with a characterization of all maximally non-classical, yet separable, $(2 \times N)$-dimensional states of rank at most two (with respect to our measure).

6.1 Introduction and results

One of the most intriguing aspects of quantum mechanics is quantum entanglement, which with the advent of quantum computing, was thrust into the limelight of quantum infor-

mation theoretic research [151]. We now know that correlations in quantum states due to entanglement are necessary in order for *pure-state* quantum computation to provide exponential speedups over its classical counterpart [160]. With bipartite entanglement nowadays fairly well understood, however, attention has turned in recent years to a more general type of quantum correlation, dubbed simply *non-classical correlations*. Unlike entanglement, such correlations *can* be created via Local Operations and Classical Communication (LOCC), but nevertheless do not exist in the classical setting. Moreover, for certain *mixed-state* quantum computational feats, the amount of entanglement present can be small or vanishing, such as in the DQC1 model of computing [174] and the locking of classical correlations [87]. In these settings, it is rather non-classical correlations which are the conjectured resource enabling such feats (see, e.g. [77, 80, 185, 78]). In fact, almost all quantum states possess non-classical correlations [94].

As a result, much attention has recently been devoted to the quantification of non-classical correlations (e.g., [187, 119, 196, 118, 217, 185, 209, 188, 13, 216, 75, 231, 208], see [195] for a survey, and Section 1.6.2 for a brief exposition). Here, we say a bipartite state ρ acting on Hilbert space $\mathcal{A} \otimes \mathcal{B}$ is *classically correlated* in $\mathcal{A}$ if and only if there exists an orthonormal basis $\{|a\rangle\}$ for $\mathcal{A}$ such that

$$\rho = \sum_i p_i |a_i\rangle\langle a_i| \otimes \rho_i \tag{6.1}$$

for $\{p_i\}$ a probability distribution and ρ_i density operators. To quantify "how far" ρ is from the form above, a number non-classicality measures, including perhaps the best-known such measure, the *quantum discord* [203, 138], ask the question of how drastically a bipartite quantum state is disturbed under local measurement on $\mathcal{A}$. In this chapter, we take a different approach to the problem. We ask: *Can disturbance of a bipartite system under local unitary operations be used to quantify non-classical correlations?*

It turns out that not only is the answer to this question *yes*, but that in fact for $(2 \times N)$-dimensional systems, the measure we construct coincides with the *geometric quantum discord* [75], a scheme based again on local measurements. Our measure is defined as follows. Given a bipartite quantum state ρ and unitary U_A acting on Hilbert spaces $\mathcal{A} \otimes \mathcal{B}$ and $\mathcal{A}$ with dimensions MN and M, respectively, define

$$D(\rho, U_A) := \frac{1}{\sqrt{2}} \left\| \rho - (U_A \otimes I_B)\, \rho \left(U_A^\dagger \otimes I_B \right) \right\|_{\mathrm{F}}, \tag{6.2}$$

where the Frobenius norm $\| A \|_{\mathrm{F}} = \sqrt{\mathrm{Tr} A^\dagger A}$ is used due to its simple calculation. Then, consider the set of unitary operators whose eigenvalues are precisely the M-th roots of

unity, i.e. whose vector of eigenvalues equals $\mathbf{v}$ for $v_k = e^{2\pi ki/M}$ for $1 \leq k \leq M$. (The corresponding eigenvectors can be chosen arbitrarily.) We call such operators *Root-of-Unity* (RU) unitaries. They include, for example, the Pauli X, Y, and Z matrices (see Section 1.4.3). Then, letting $\mathrm{RU}(\mathcal{A})$ denote the set of RU unitaries acting on $\mathcal{A}$, we define our measure as:

$$D(\rho) := \min_{U_A \in \mathrm{RU}(\mathcal{A})} D(\rho, U_A). \tag{6.3}$$

Note that $0 \leq D(\rho) \leq 1$ for all ρ acting on $\mathcal{A} \otimes \mathcal{B}$.

Our results: In this chapter, we show the following regarding $D(\rho)$.

1. Closed form expressions. Our first result is a closed-form expression for $D(\rho)$ for $(2 \times N)$-dimensional systems (Theorem 6.3). This reveals that for $(2 \times N)$-dimensional ρ, $D(\rho)$ coincides with the geometric discord of ρ. It also allows us to prove that, like the *Fu distance* [102, 107] (defined below in *Previous Work*), if $D(\rho) > 1/\sqrt{2}$, then ρ violates the Clauser-Horne-Shimony-Holt (CHSH) inequality [70] (Corollary 6.5). We also derive a closed form expression for $D(\rho)$ for Werner states, finding here that $D(\rho)$ in fact equals the Fu distance of ρ (Theorem 6.6).

2. States achieving $D(\rho) = 1$. We next show that only pure maximally entangled states ρ achieve the maximum value $D(\rho) = 1$, as expected (Corollary 6.8).

3. $D(\rho)$ is faithful. We show that $D(\rho)$ is a *faithful* non-classicality measure, i.e. it achieves a value of zero if and only if ρ is classically correlated in $\mathcal{A}$ (Theorem 6.10). To prove this, we first derive a new characterization of states with zero quantum discord based on the correlation matrix of ρ. We then show that the states achieving $D(\rho) = 0$ can be characterized in the same way. More generally, by extending our scheme to allow the eigenvalues of U_A to have multiplicity at most k, we prove a state is undisturbed under U_A if and only if there exists a projective measurement on $\mathcal{A}$ of rank at most k acting invariantly on the state (Theorem 6.11). This reproduces in a simple fashion a result of Reference [197] regarding entanglement quantification in the pure state setting. Based on this equivalence between disturbance under local unitary operations and local projective measurements, we propose a generalized definition of the quantum discord at the end of Section 6.6.

4. Maximally non-classical, yet separable states. Finally, we characterize the set of maximally non-classical, yet separable, $(2 \times N)$-dimensional ρ of rank at most two, according to $D(\rho)$ (and hence according to the geometric discord) (Lemmas 6.13 and 6.14).

Previous work: The Fu distance, defined as the *maximization* of Equation (6.2) over all U_A such that $[U_A, \mathrm{Tr}_B(\rho)] = 0$, was defined in Reference [102] and studied further in References [107] and [78] with regards to quantifying entanglement and non-classicality. Despite its strengths, such as a closed form solution for two-qubit systems and Werner states, and a connection to the CHSH inequality, the distance has weaknesses: It can attain its maximum value even on non-maximally entangled pure states [107], and is not a faithful non-classicality measure [78]. Interestingly, our $D(\rho)$ eliminates these weaknesses while preserving the former strengths. Subsequent to the conception of our scheme, the present author learned that there has also been an excellent line of work studying (the square of) Equation (6.3) in another setting — that of *pure state entanglement*. In Reference [112], it was found that in $(2 \times N)$ and $(3 \times N)$ systems, $D(|\psi\rangle\langle\psi|)^2$ coincides with the *linear entropy of entanglement*. Reference [197] then showed that for arbitrary bipartite pure states, $D(|\psi\rangle\langle\psi|)^2$ is a faithful entanglement monotone, and derived upper and lower bounds in terms of the linear entropy of entanglement. Finally, alternative characterizations of zero discord states have been given in [203, 75, 76]. Maximally non-classical separable two-qubit states have been studied, for example, in [110, 114]. For example, the set of such states found [110] with respect to the *relative entropy of quantumness* matches our characterization for $D(\rho)$; we remark, however, that our analysis for $D(\rho)$ in this regard is more general than in [110] as it is based on a less restrictive ansatz. We remark that since the initial posting of the paper this chapter is based on, a related work by Streltsov *et al.* has appeared [229].

Discussion and open questions: Our results show that local unitary operations can indeed form the basis of a non-classicality measure with certain desirable properties. In particular, the scheme we consider is faithful, correctly identifies maximally non-classical states, and reveals interesting connections to a number of quantifiers of correlations, such as the Fu distance, the quantum discord, the geometric quantum discord, and the relative entropy of quantumness. As outlined above, the strengths of our scheme include a closed form for two-qubit states and Werner states, the former of which reveals a link between the paradigms of "disturbance under local unitary operations" and "disturbance under local measurements" by reducing to the geometric discord for two-qubit states. This link is further strengthened by the demonstration of connections to even generalized versions of the quantum discord.

We leave open the following questions. For what other interesting classes of quantum states can a closed form expression for $D(\rho)$ be found? Can a better intuitive understanding of the interplay between the notions of "disturbance under local measurements" and

"disturbance under local unitary operations" be obtained in higher dimensions? We give an analytical characterization of all maximally non-classical rank-two $(2 \times N)$-dimensional separable states — we conjecture that higher rank two-qubit states, for example, achieve strictly smaller values of $D(\rho)$. Can this be proven rigorously and analytically? (We remark that a numerical proof for this conjecture was given in [114] for the geometric discord, for example.) What can the study of the generalized notion of quantum discord we define in Section 6.6, $\delta_{\mathbf{v}}(\rho)$, tell us about non-classical correlations?

Organization of this chapter: We begin in Section 6.2 with necessary definitions and useful lemmas. Closed forms for $(2 \times N)$-dimensional systems are given in Section 6.3 and for Werner states in Section 6.4. Section 6.5 characterizes the set of states achieving $D(\rho) = 1$. Section 6.6 shows that $D(\rho)$ is faithful. In Section 6.7, we discuss maximally non-classical separable states.

6.2 Preliminaries

We begin by reviewing notation specific to this chapter, followed by relevant definitions and useful lemmas. Throughout this chapter, we use $\mathcal{A}$ and $\mathcal{B}$ to denote complex Euclidean spaces of dimensions M and N, respectively. We define $\rho_A := \mathrm{Tr}_B(\rho)$ and $\rho_B := \mathrm{Tr}_A(\rho)$. The anti-commutator of A and B is $\{A, B\} = AB + BA$. The notation $\mathrm{diag}(\mathbf{v})$ for complex vector $\mathbf{v}$ denotes a diagonal matrix with ith diagonal entry v_i, and $\mathrm{span}(\{\mathbf{v}_i\})$ denotes the span of the set of vectors $\{\mathbf{v}_i\}$.

Moving to definitions, in this chapter we often decompose $\rho \in \mathcal{D}(\mathcal{A} \otimes \mathcal{B})$ in terms of a Hermitian basis for $\mathcal{H}(\mathcal{A} \otimes \mathcal{B})$ (sometimes known as the Fano form [91]):

$$\begin{aligned} \rho = \tfrac{1}{MN} \quad & (I^A \otimes I^B + \mathbf{r}^A \cdot \sigma^A \otimes I^B + \\ & I^A \otimes \mathbf{r}^B \cdot \sigma^B + \sum_{i=1}^{M^2-1} \sum_{j=1}^{N^2-1} T_{ij} \sigma_i^A \otimes \sigma_j^B). \end{aligned} \tag{6.4}$$

Here, σ^A is a $(M^2 - 1)$-component vector of traceless orthogonal Hermitian basis elements σ_i^A satisfying $\mathrm{Tr}(\sigma_i^A \sigma_j^A) = 2\delta_{ij}$, $\mathbf{r}^A \in \mathbb{R}^{M^2-1}$ is the Bloch vector for subsystem A with $r_i^A = \frac{M}{2}\mathrm{Tr}(\rho_A \sigma_i^A)$, and $T \in \mathbb{R}^{(M^2-1)\times(N^2-1)}$ is the *correlation matrix* with entries $T_{ij} = \frac{MN}{4}\mathrm{Tr}(\sigma_i^A \otimes \sigma_j^B \rho)$. For $M = 2$, $\mathbf{r}_A$ satisfies $0 \leq \| \mathbf{r}_A \|_2 \leq 1$ with $\| \mathbf{r}_A \|_2 = 1$ if and only if ρ_A is pure. The definitions for subsystem B are analogous.

We now give a useful specific construction for the basis elements σ_i^A [139]. Define $\{\sigma_i\}_{i=1}^{M^2-1} = \{U_{pq}, V_{pq}, W_r\}$, such that for $1 \leq p < q \leq M$ and $1 \leq r \leq M-1$, and $\{|i\rangle\}_{i=1}^{M}$ some orthonormal basis for $\mathcal{A}$:

$$U_{pq} = |p\rangle\langle q| + |q\rangle\langle p| \tag{6.5}$$

$$V_{pq} = -i|p\rangle\langle q| + i|q\rangle\langle p| \tag{6.6}$$

$$W_r = \sqrt{\frac{2}{r(r+1)}}\left(\sum_{k=1}^{r} |k\rangle\langle k| - r|r+1\rangle\langle r+1|\right). \tag{6.7}$$

Note that when $M = 2$, this construction yields the Pauli matrices $\sigma^{\mathbf{A}} = (X, Y, Z)$.

Regarding $D(\rho)$, defining $\rho_f := (U_A \otimes I_B)\rho(U_A^\dagger \otimes I_B)$, we often use the fact that Equation (6.3) can be rewritten as:

$$D(\rho) = \min_{U_A \in \mathrm{RU}(\mathcal{A})} \sqrt{\mathrm{Tr}(\rho^2) - \mathrm{Tr}(\rho\rho_f)}. \tag{6.8}$$

Finally, we show a simple but important lemma.

Lemma 6.1. *$D(\rho)$ is invariant under local unitary operations.*

Proof. Let $\rho' := (V_A \otimes V_B)\rho(V_A \otimes V_B)^\dagger$ for unitaries V_A, V_B. Then in Equation (6.8), $\mathrm{Tr}(\rho'^2) = \mathrm{Tr}(\rho^2)$, and $\mathrm{Tr}(\rho'\rho'_f)$ becomes

$$\mathrm{Tr}(\rho(V_A^\dagger U_A V_A \otimes I_B)\rho(V_A^\dagger U_A^\dagger V_A \otimes I_B)). \tag{6.9}$$

Observe, however, that $V_A U_A V_A^\dagger$ is still an RU unitary, since we have simply changed basis. Hence, $D(\rho', U_A) = D(\rho, V_A^\dagger U_A V_A)$, and since we are minimizing over all $U_A \in \mathrm{RU}(\mathcal{A})$, the claim follows. □

6.3 $(2 \times N)$-dimensional states

In this section, we study $D(\rho)$ for $\rho \in \mathcal{D}(\mathbb{C}^2 \otimes \mathbb{C}^N)$, obtaining among other results a closed from expression for $D(\rho)$. To begin, note that any $U_A \in \mathrm{RU}(\mathcal{A})$ must have the form

$$U_A := |c\rangle\langle c| - |d\rangle\langle d| = 2|c\rangle\langle c| - I_2, \tag{6.10}$$

up to an irrelevant global phase which disappears upon application of U_A to our system, and for some orthonormal basis $\{|c\rangle, |d\rangle\}$ for $\mathbb{C}^2$. Then, $D(\rho, U_A)$ can be rewritten as

$$2\sqrt{\mathrm{Tr}[\rho^2(|c\rangle\langle c| \otimes I) - \rho(|c\rangle\langle c| \otimes I)\rho(|c\rangle\langle c| \otimes I)]}. \tag{6.11}$$

We begin with a simple upper bound on $D(\rho)$.

Theorem 6.2. *For any $\rho \in \mathcal{D}(\mathbb{C}^2 \otimes \mathbb{C}^N)$, one has*

$$D(\rho) \leq 2\sqrt{\lambda_{\min}(\mathrm{Tr}_{\mathcal{B}}(\rho^2))}. \tag{6.12}$$

Proof. Starting with Equation (6.11), by noting that $\mathrm{Tr}[\rho(|c\rangle\langle c| \otimes I)\rho(|c\rangle\langle c| \otimes I)] \geq 0$ and using the fact that $\mathrm{Tr}(\rho(C_A \otimes I_B)) = \mathrm{Tr}(\rho_A C_A)$, we have that $D(\rho)$ is at most

$$\min_{\text{unit } |c\rangle \in \mathbb{C}^2} 2\sqrt{\mathrm{Tr}[\mathrm{Tr}_{\mathcal{B}}(\rho^2)|c\rangle\langle c|]} = 2\sqrt{\lambda_{\min}(\mathrm{Tr}_{\mathcal{B}}(\rho^2))}. \quad \square \tag{6.13}$$

Theorem 6.2 implies that for pure product $|\psi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^N$, $D(|\psi\rangle\langle\psi|) = 0$, in agreement with the results in Reference [112]. By next exploiting the structure of ρ further, we obtain a closed form expression for $D(\rho)$.

Theorem 6.3. *For any $\rho \in \mathcal{D}(\mathbb{C}^2 \otimes \mathbb{C}^N)$, define $G := \mathbf{r}^A(\mathbf{r}^A)^T + \frac{2}{N}TT^T$, for T the correlation matrix of ρ. Then, $D(\rho)$ equals*

$$\frac{1}{\sqrt{N}}\sqrt{\mathrm{Tr}(G) - \lambda_{\max}(G)} = \frac{1}{\sqrt{N}}\sqrt{\lambda_2(G) + \lambda_3(G)}. \tag{6.14}$$

Proof. Define $P := |c\rangle\langle c|$. Then, beginning with Equation (6.11), by rewriting ρ using Equation (6.4) and applying the fact that the basis elements σ_i are traceless, we obtain that $\mathrm{Tr}(\rho^2 P \otimes I - \rho P \otimes I \rho P \otimes I)$ equals

$$\frac{1}{4N}\mathrm{Tr}(A_1 - A_2 + A_3 - A_4), \tag{6.15}$$

where

$$A_1 := \left(\sum_i r_i^A \sigma_i{}^A\right)^2 P, \qquad A_2 := \left(\sum_i r_i^A \sigma_i{}^A P\right)^2 \tag{6.16}$$

$$A_3 := \frac{1}{N}\left(\sum_{ij} T_{ij}\sigma_i^A \otimes \sigma_j^B\right)^2 (P \otimes I) \tag{6.17}$$

$$A_4 := \frac{1}{N}\left(\sum_{ij} T_{ij}\sigma_i^A \otimes \sigma_j^B\right)\left(\sum_{ij} T_{ij} P\sigma_i^A P \otimes \sigma_j^B\right). \tag{6.18}$$

Using the facts that $(\sigma_i^A)^2 = I$, $\{\sigma_i^A, \sigma_j^A\} = 0$ for $i \neq j$, $\mathrm{Tr}(\sigma_i\sigma_j) = 2\delta_{ij}$, and $\mathrm{Tr}(P) = 1$, we thus have

$$\mathrm{Tr}(A_1) = \| \mathbf{r}^A \|_2^2, \qquad \mathrm{Tr}(A_3) = \frac{2}{N}\sum_{ij} T_{ij}^2 \tag{6.19}$$

$$\mathrm{Tr}(A_2) = \sum_{ij} r_i^A r_j^A \langle c|\sigma_i^A|c\rangle\langle c|\sigma_j^A|c\rangle \tag{6.20}$$

$$\mathrm{Tr}(A_4) = \frac{2}{N}\sum_{ij}\left(\sum_k T_{ik}T_{jk}\right)\langle c|\sigma_i^A|c\rangle\langle c|\sigma_j^A|c\rangle. \tag{6.21}$$

Now, $\langle c|\sigma_i^A|c\rangle$ can be thought of as the ith component of the Bloch vector $\mathbf{v} \in \mathbb{R}^3$ of pure state $|c\rangle$ with $\| \mathbf{v} \|_2 = 1$, implying

$$\mathrm{Tr}(A_2 + A_4) = \mathbf{v}^T\left[\mathbf{r}^A(\mathbf{r}^A)^T + \frac{2}{N}TT^T\right]\mathbf{v}. \tag{6.22}$$

Plugging these values into Equation (6.11), we conclude $D(\rho)$ equals

$$\min_{\substack{\mathbf{v}\in\mathbb{R}^3 \\ \| \mathbf{v} \|_2 = 1}} \frac{1}{\sqrt{N}}\sqrt{\| \mathbf{r}^A \|_2^2 + \frac{2}{N}\sum_{ij} T_{ij}^2 - \mathrm{Tr}(A_2 + A_4)}. \tag{6.23}$$

The claim now follows since for any symmetric $A \in \mathbb{R}^{n\times n}$, $\max_{\text{unit } \mathbf{v}\in\mathbb{R}^n} \mathbf{v}^T A\mathbf{v} = \lambda_{\max}(A)$. □

The expression for $D(\rho)$ in Theorem 6.3 matches that for the *geometric discord* [75, 242]. Specifically, defining the latter as $\delta_g(\rho) = \min_{\sigma\in\Omega}\sqrt{2}\,\| \rho - \sigma \|_{\mathrm{F}}$, where Ω is the set of zero-discord states, we have for $(2 \times N)$-dimensional ρ that $D(\rho) = \delta_g(\rho)$. (Note: The original definition of Reference [75] was more precisely $\delta_g(\rho) = \min_{\sigma\in\Omega} \| \rho - \sigma \|_{\mathrm{F}}^2$.)

We now discuss consequences of Theorem 6.3, beginning with a lower bound which proves useful later.

Corollary 6.4. *For $\rho \in \mathcal{D}(\mathbb{C}^2 \otimes \mathbb{C}^N)$, we have*

$$D(\rho) \geq \frac{\sqrt{2}}{N}\sqrt{\lambda_2(TT^T) + \lambda_3(TT^T)}. \tag{6.24}$$

This holds with equality if $\mathbf{r}^A = 0$, i.e. $\rho_A = \frac{I}{2}$.

Proof. The first claim follows from the fact that:

$$\lambda_{\max}\left(\mathbf{r}^A(\mathbf{r}^A)^T + \frac{2}{N}TT^T\right) \leq \left\|\,\mathbf{r}^A\,\right\|_2^2 + \frac{2}{N}\lambda_{\max}\left(TT^T\right). \tag{6.25}$$

The second claim follows by substitution into Equation (6.14). □

For example, for maximally entangled $|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, for which $\mathbf{r}^B = \mathbf{0}$ and $T = \text{diag}(1, -1, 1)$, Corollary 6.4 yields $D(|\psi\rangle\langle\psi|) = 1$, as desired. We also remark that Equation (6.14) can further be simplified for two-qubit states, since by Reference [148, 149], one can assume without loss of generality that T is diagonal. This relies on the facts that (1) applying local unitary $V_1 \otimes V_2$ to ρ has the effect of mapping $T \mapsto O_1 T O_2^\dagger$, $\mathbf{r}^A \mapsto O_1\mathbf{r}^A$, and $\mathbf{r}^B \mapsto O_2\mathbf{r}^B$ for some orthogonal rotation matrices O_1 and O_2, and (2) $D(\rho)$ is invariant under local unitaries by Lemma 6.1.

Using Corollary 6.4, we next obtain a connection to the CHSH inequality for two-qubit ρ. Defining $M(\rho) := \lambda_1(T^TT) + \lambda_2(T^TT)$, it is known that ρ violates the CHSH inequality if and only if $M(\rho) > 1$ [150]. We thus have:

Corollary 6.5. *For $\rho \in \mathcal{D}(\mathbb{C}^2 \otimes \mathbb{C}^2)$, if $D(\rho) > 1/\sqrt{2}$, then $M(\rho) > 1$. The converse does not hold.*

Proof. The first is immediate from Corollary 6.4 and the fact that TT^T and T^TT are cospectral (Theorem 1.3.20 of [143]). The converse proceeds similarly to Theorem 7 of Reference [107] — namely, let $|\psi\rangle = a|00\rangle + b|11\rangle$ for real $a, b \geq 0$ and $a^2 + b^2 = 1$. Then, for density operator $|\psi\rangle\langle\psi|$, we have $\mathbf{r}^B = (0, 0, a^2 - b^2)$ and $T = \text{diag}(2ab, -2ab, 1)$, implying $M(|\psi\rangle\langle\psi|) > 1$ for $a, b \neq 0$. In comparison, $D(|\psi\rangle\langle\psi|) = 2ab \leq 1/\sqrt{2}$ when $a \leq \sqrt{\frac{1}{2} - \frac{1}{2\sqrt{2}}}$ or $a \geq \sqrt{\frac{1}{2} + \frac{1}{2\sqrt{2}}}$. □

Interestingly, the exact same relationship as that in Corollary 6.5 was found between the Fu distance and the CHSH inequality in Reference [107].

6.4 Werner states

We now derive a closed formula for $D(\rho)$ for Werner states $\rho \in \mathcal{D}(\mathbb{C}^d \otimes \mathbb{C}^d)$ where $d \geq 2$, which are defined as [254]

$$\rho := \frac{2p}{d^2 + d}P_s + \frac{2(1-p)}{d^2 - d}P_a, \tag{6.26}$$

for $P_s := (I+P)/2$ and $P_a := (I-P)/2$ the projectors onto the symmetric and anti-symmetric subspaces, respectively, $P := \sum_{i,j=1}^{d} |i\rangle\langle j| \otimes |j\rangle\langle i|$ the SWAP operator, and $0 \leq p \leq 1$. Werner states are invariant under $U \otimes U$ for any unitary U, and are entangled if and only if $p < 1/2$.

Theorem 6.6. *Let $\rho \in \mathcal{D}(\mathbb{C}^d \otimes \mathbb{C}^d)$ be a Werner state. Then*

$$D(\rho) = \frac{|2pd - d - 1|}{d^2 - 1}. \tag{6.27}$$

Proof. As done in Theorem 3 of Reference [107], we first rewrite Equation 6.8 using the facts that $\mathrm{Tr}(P) = d$, $\mathrm{Tr}(P^2) = d^2$, and $\beta := \mathrm{Tr}(P(U_A \otimes I)P(U_A \otimes I)^\dagger) = \mathrm{Tr}(U_A)\mathrm{Tr}(U_A^\dagger)$ to obtain that for any $U_A \in U(\mathcal{A})$,

$$D(\rho, U_A) = \frac{\sqrt{(2pd - d - 1)^2(d^2 - \beta)}}{d(d^2 - 1)}. \tag{6.28}$$

Since $\mathrm{Tr}(U_A) = 0$ for any $U_A \in \mathrm{RU}(\mathcal{A})$, we have $\beta = 0$ and the claim follows. □

Again, we find that this coincides exactly with the expression for the Fu distance for Werner states [107]. Further, Theorem 6.6 implies that the quantum discord of Werner state ρ is zero if and only if $p = (d+1)/2d$. This matches the results of Chitambar [67], who develops the following closed formula for the discord $\delta(\rho)$ of Werner states:

$$\begin{aligned} \delta(\rho) \quad = \quad & \log(d+1) + (1-p)\log\frac{1-p}{d-1} + p\log\frac{p}{d+1} - \\ & \frac{2p}{d+1}\log p - \left(1 - \frac{2p}{d+1}\right)\log\frac{d+1-2p}{2(d-1)}. \end{aligned} \tag{6.29}$$

In Section 6.6, we show that this is no coincidence — it turns out that $D(\rho) = 0$ if and only if the discord of ρ is zero for any ρ.

6.5 Pure states of arbitrary dimension

We now show that only pure maximally entangled states ρ achieve $D(\rho) = 1$. As mentioned in Section 6.1, this is in contrast to the Fu distance [102, 107], whose maximal value is attained even for certain *non-maximally* entangled $|\psi\rangle$. We remark that Theorem 6.7 below also follows from a more general non-trivial result that $D(|\psi\rangle\langle\psi|)^2$ is tightly upper

bounded by the linear entropy of entanglement of pure state $|\psi\rangle$ [197]. However, our proof of Theorem 6.7 is much simpler and requires only elementary linear algebra.

To begin, assume without loss of generality that $M \leq N$, and let $|\psi\rangle \in \mathcal{A} \otimes \mathcal{B}$ be a pure quantum state with Schmidt decomposition $|\psi\rangle = \sum_{k=1}^{M} \alpha_k |a_k\rangle \otimes |b_k\rangle$, i.e. $\sum_k \alpha_k^2 = 1$ for $\alpha_k \in \mathbb{R}$ and $\{|a_k\rangle\}$ and $\{|b_k\rangle\}$ the Schmidt bases for $\mathcal{A}$ and $\mathcal{B}$, respectively.

Theorem 6.7. *Let $|\psi\rangle \in \mathcal{A} \otimes \mathcal{B}$ with Schmidt decomposition as above. Then $D(|\psi\rangle\langle\psi|) = 1$ if and only if $\alpha_k = \frac{1}{\sqrt{M}}$ for all $1 \leq k \leq M$ (i.e. $|\psi\rangle$ is maximally entangled).*

Proof. We begin by rewriting Equation (6.8) as

$$D(|\psi\rangle\langle\psi|) = \min_{U_A \in \mathrm{RU}(\mathcal{A})} \sqrt{1 - \left| \sum_{k=1}^{M} \alpha_k^2 \langle a_k | U_A | a_k \rangle \right|^2}. \tag{6.30}$$

If $|\psi\rangle$ is maximally entangled, then $\alpha_k = 1/\sqrt{M}$ for all $1 \leq k \leq M$. Then, since $U_A \in \mathrm{RU}(\mathcal{A})$, Equation (6.30) yields

$$D(|\psi\rangle\langle\psi|) = \min_{U_A \in \mathrm{RU}(\mathcal{A})} \sqrt{1 - \frac{1}{M^2} \left|\mathrm{Tr}(U_A)\right|^2} = 1. \tag{6.31}$$

For the converse, assume $D(|\psi\rangle\langle\psi|) = 1$. Then, by Equation (6.30), we must have that for all $U_A \in \mathrm{RU}(\mathcal{A})$,

$$\sum_{k=1}^{M} \alpha_k^2 \langle a_k | U_A | a_k \rangle = 0. \tag{6.32}$$

Thus, choosing U_A as diagonal in basis $\{|a_k\rangle\}$, Equation (6.32) says that $\mathbf{w}^T \pi \mathbf{v} = 0$ for all permutations $\pi \in S_M$, where $w_k := \alpha_k^2$ and $v_k := e^{2\pi k i / M}$. This can only hold, however, if all entries of $\mathbf{w}$ are the same, i.e. $\alpha_k = 1/\sqrt{M}$ for all $1 \leq k \leq M$, as desired. □

Corollary 6.8. *A quantum state $\rho \in \mathcal{D}(\mathcal{A} \otimes \mathcal{B})$ achieves $D(\rho) = 1$ if and only if ρ is pure and maximally entangled.*

Proof. Immediate from Theorem 6.7 and the $\mathrm{Tr}(\rho^2)$ in Equation (6.8). □

6.6 Relationship to quantum discord

We now show that for arbitrary $\rho \in \mathcal{D}(\mathcal{A} \otimes \mathcal{B})$, $D(\rho)$ is zero if and only if the quantum discord [203, 138] $\delta(\rho)$ of ρ is zero. (The discord $\delta(\rho)$ was defined in Section 1.6.2.)

The main fact we leverage about the discord here is the following.

Theorem 6.9 (Ollivier and Zurek [203])**.** *For $\rho \in \mathcal{D}(\mathcal{A} \otimes \mathcal{B})$, $\delta(\rho) = 0$ if and only if*

$$\rho = \sum_j \Pi_j^A \otimes I^B \rho \Pi_j^A \otimes I^B, \tag{6.33}$$

for some complete set of rank 1 projectors $\left\{\Pi_j^A\right\}$.

We now prove the main result of this section. The first part of the proof involves a new characterization of the set of zero discord quantum states ρ in terms of the basis elements σ_i^A from the Fano form of ρ. Key to this characterization is the absence of non-diagonal σ_i^A in the expansion of ρ. In the proofs below, we assume the basis elements σ_i^A for $\mathcal{A}$ come from the set $\{I, U_{pq}, V_{pq}, W_r\}_{p,q,r}^A$ from Section 6.2 (analogously for $\mathcal{B}$).

Theorem 6.10. *Let $\rho \in \mathcal{D}(\mathcal{A} \otimes \mathcal{B})$. Then $\delta(\rho) = 0$ if and only if there exists a local unitary V^A such that*

$$\operatorname{Tr}\left(\left(V^A \otimes I^B\right) \rho \left(V^{A\dagger} \otimes I^B\right) \left(\sigma_i^A \otimes \sigma_j^B\right)\right) = 0 \tag{6.34}$$

for all $\sigma_i^A \in \{U_{pq}, V_{pq}\}^A$ and all $\sigma_j^B \in \{I, U_{pq}, V_{pq}, W_r\}^B$. The same characterization holds for $D(\rho) = 0$.

Proof. We prove the equivalent statement that $\delta(\rho) = 0$ if and only if there exists an orthonormal basis $\{|k\rangle\}$ for $\mathcal{A}$ such that, for basis elements σ_i^A constructed with respect to $\{|k\rangle\}$, we have $\operatorname{Tr}(\rho(\sigma_i^A \otimes \sigma_j^B)) = 0$ for all $\sigma_i^A \in \{U_{pq}, V_{pq}\}$ (and similarly for $D(\rho) = 0$).

Suppose $\delta(\rho) = 0$. Then by Theorem 6.9, there exists a complete set of rank 1 projectors $\left\{\Pi_j^A\right\}$ such that Equation (6.33) holds. Let $\{|k\rangle\}$ be the basis onto which $\left\{\Pi_j^A\right\}$ projects, and define $\Phi(C) := \sum_j \Pi_j^A C \Pi_j^A$. By constructing the basis elements σ_i^A in Equation (6.4) using $\{|k\rangle\}$, we thus have

$$\begin{aligned} \rho \;=\; & \frac{1}{MN}\Big[I^A \otimes I^B + I^A \otimes \mathbf{r}^B \cdot \sigma^B + \\ & \sum_{i=1}^{M^2-1} \Phi(\sigma_i^A) \otimes \left(r_i^A I^B + \sum_{j=1}^{N^2-1} T_{ij} \sigma_j^B\right)\Big]. \end{aligned} \tag{6.35}$$

Now, for all $\sigma_i^A \in \{W_r\}$, we clearly have $\Phi(\sigma_i^A) = \sigma_i^A$. For $\sigma_i^A \in \{U_{pq}, V_{pq}\}$, however, $\Phi(\sigma_i^A) = 0$. Thus, in order for Equation (6.33) to hold, we must have $r_i^A = T_{ij} = 0$ for all basis elements $\sigma_i^A \in \{U_{pq}, V_{pq}\}$, which by definition means $\mathrm{Tr}(\rho(\sigma_i^A \otimes \sigma_j^B)) = 0$ for all $\sigma_i^A \in \{U_{pq}, V_{pq}\}^A$, as desired. To show that this implies $D(\rho) = 0$, construct $U^A \in \mathrm{RU}(\mathcal{A})$ as diagonal in basis $\{|k\rangle\}$ and define $\Phi(C) := U^A C U^{A\dagger}$. Then since in Equation (6.35), we have $\Phi(\sigma_i^A) = \sigma_i^A$ for any $\sigma_i^A \in \{I, W_r\}$, the claim follows.

To show the converse, assume $D(\rho, U^A) = 0$ for some $U^A \in \mathrm{RU}(\mathcal{A})$. Then, construct the basis elements σ_i^A with respect to a diagonalizing basis $\{|k\rangle\}$ for U^A and define $\Phi(C) := U^A C U^{A\dagger}$. It follows that for any p and q,

$$\Phi(U_{pq}) = e^{i(\theta_p - \theta_q)}|p\rangle\langle q| + e^{-i(\theta_p-\theta_q)}|q\rangle\langle p|, \tag{6.36}$$

$$\Phi(V_{pq}) = -ie^{i(\theta_p - \theta_q)}|p\rangle\langle q| + ie^{-i(\theta_p-\theta_q)}|q\rangle\langle p|. \tag{6.37}$$

Consider now an arbitrary term $(c_u\sigma_u^A + c_v\sigma_v^A)\otimes\sigma_j^B$ from the Fano form of ρ where $\sigma_u^A = U_{pq}$ and $\sigma_v^B = V_{pq}$ for some choice of p and q. Since Equations (6.36) and (6.37) imply that U^A can only map U_{pq} to V_{pq} and vice versa, it follows that in order for $D(\rho, U^A) = 0$ to hold, we must have $\Phi(c_u\sigma_u^A + c_v\sigma_v^A) = c_u\sigma_u^A + c_v\sigma_v^A$. This leads to the system of equations

$$c_u - ic_v = e^{i(\theta_p - \theta_q)}(c_u - ic_v) \tag{6.38}$$

$$c_u + ic_v = e^{-i(\theta_p - \theta_q)}(c_u + ic_v). \tag{6.39}$$

We conclude that if either $c_u \neq 0$ or $c_v \neq 0$, it must be that $\theta_p = \theta_q$ in order for $D(\rho) = 0$ to hold. However, since all eigenvalues of U^A are distinct by definition, this is impossible. Thus, $\mathrm{Tr}(\rho(\sigma_i^A \otimes \sigma_j^B)) = 0$ for all $\sigma_i^A \in \{U_{pq}, V_{pq}\}$, as desired. To see that this implies $\delta(\rho) = 0$, simply now choose $\left\{\Pi_j^A\right\}$ as the projection onto $\{|k\rangle\}$. Then, defining $\Phi(C) := \sum_j \Pi_j^A C \Pi_j^A$ and applying the same arguments from the forward direction to Equation (6.35), we conclude that ρ is invariant under $\left\{\Pi_j^A\right\}$. By Theorem 6.9, we have $\delta(\rho) = 0$, completing the proof. □

Theorem 6.10 shows that $D(\rho)$ defined in Equation (6.3) is zero precisely for the set of states classically correlated in $\mathcal{A}$. In other words, unlike the Fu distance [78], $D(\rho)$ is indeed a *faithful* non-classicality measure. The proof of Theorem 6.10 does, however, have a curiosity — the key property the proof relies on is that all $U^A \in \mathrm{RU}(\mathcal{A})$ have non-degenerate spectra. Interestingly, this is the mixed-state analogue of the pure-state result of Reference [197], where it was shown that a non-degenerate spectrum suffices to conclude $D(|\psi\rangle\langle\psi|)$ is a faithful entanglement monotone for pure states $|\psi\rangle$. Specifically, Reference [197] shows that if in Equation (6.3) we minimize over U^A with eigenvalues of

multiplicity at most k (with at least one eigenvalue of multiplicity k), then $D(|\psi\rangle\langle\psi|) = 0$ if and only if $|\psi\rangle$ has Schmidt rank at most k. Could there be an analogue of this more general result in the mixed-state setting of non-classicality? It turns out the answer is *yes*.

Let $\mathbf{v} \in \mathbb{N}^M$ such that $\sum_{j=1}^M v_j j = M$. Then, consider an arbitrary (i.e. not necessarily RU) unitary $U_{\mathbf{v}}^A$ which has precisely v_j distinct eigenvalues with multiplicity j. For example, $U_{\mathbf{v}}^A \in \mathrm{RU}(\mathcal{A})$ has $\mathbf{v} = (M, 0, \ldots, 0)$ since it has M distinct eigenvalues of multiplicity 1. Similarly, if $\mathbf{v} = (0, 0, \ldots, 1)$, then $U_{\mathbf{v}}^A$ is just the identity (up to phase), and if $\mathbf{v} = (M-4, 2, \ldots, 0)$ then $U_{\mathbf{v}}^A$ has $M-4$ distinct eigenvalues of multiplicity 1, and two distinct eigenvalues with multiplicity 2 each. Now, corresponding to any $U_{\mathbf{v}}^A$ is a complete projective measurement $\left\{\Pi_j^A\right\}_{\mathbf{v}}$ which consists precisely of v_j projectors of rank j. The correspondence is simple: Let λ be an eigenvalue of $U_{\mathbf{v}}^A$ with multiplicity j, i.e. the projector Π_λ onto its eigenspace has rank j. Then $\Pi_\lambda \in \left\{\Pi_j^A\right\}_{\mathbf{v}}$. It is easy to see that similarly, corresponding to any $\left\{\Pi_j^A\right\}_{\mathbf{v}}$ is a $U_{\mathbf{v}}^A$ (assuming we are not concerned with the precise eigenvalues of $U_{\mathbf{v}}^A$, as is this case here). We can now state the following.

Theorem 6.11. *Let $\rho \in \mathcal{D}(\mathcal{A} \otimes \mathcal{B})$ and $\mathbf{v} \in \mathbb{N}^M$ such that $\sum_{j=1}^M v_j j = M$. Then, there exists a complete projective measurement $\left\{\Pi_j^A\right\}_{\mathbf{v}}$ such that*

$$\rho = \sum_j \Pi_j^A \otimes I^B \rho \Pi_j^A \otimes I^B \tag{6.40}$$

if and only if there exists a $U_{\mathbf{v}}^A \in \mathcal{U}(\mathcal{A})$ with $D(\rho, U_{\mathbf{v}}^A) = 0$.

Proof. The proof follows that of Theorem 6.10, so we outline the differences. Here, $U_{\mathbf{v}}^A$ and $\left\{\Pi_j^A\right\}_{\mathbf{v}}$ will be related through the correspondence outlined above, and the basis elements σ_i^A are constructed with respect to a diagonalizing basis $\{|k\rangle\}$ for $U_{\mathbf{v}}^A$ (which by definition also diagonalizes each $\Pi_j^A \in \left\{\Pi_j^A\right\}_{\mathbf{v}}$). For simplicity, we discuss the case of $\mathbf{v} = (M-2, 1, 0, \ldots, 0)$; all other cases proceed analogously.

Going in the forward direction, suppose $\Pi_j^A \in \left\{\Pi_j^A\right\}_{\mathbf{v}}$ projects onto $\mathcal{S}_{pq} := \mathrm{span}(|p\rangle, |q\rangle)$. Then, in Equation (6.35), $\Phi(\sigma_i^A) = \sigma_i^A$ for $\sigma_i^A = U_{pq}$ and $\sigma_i^A = V_{pq}$. In other words, now we can have $r_i^A \neq 0$ and $T_{ij} \neq 0$ (however, note we still have $r_{m\neq i}^A = 0$ and $T_{m\neq i,j} = 0$). Since $U_{\mathbf{v}}^A$ has a degenerate eigenvalue on S_{pq}, however, we have by Equations (6.36) and (6.37) that $U_{\mathbf{v}}^A$ acts invariantly on σ_i^A as well (since $\theta_p = \theta_q$). The converse is similar; namely, suppose $U_{\mathbf{v}}^A$ has a degenerate eigenvalue on $\mathcal{S}_{pq}$. Then the projector onto the corresponding two-dimensional eigenspace $\Pi_j^A \in \left\{\Pi_j^A\right\}_{\mathbf{v}}$ is $\Pi_j^A = |p\rangle\langle p| + |q\rangle\langle q|$. It thus follows by the same argument as above that both $U_{\mathbf{v}}^A$ and Π_j^A act invariantly on U_{pq} and V_{pq}. □

From this general theorem, we can re-derive as a simple corollary the pure state result of Reference [197] mentioned earlier, which we rephrase in our terminology as follows.

Corollary 6.12. *Let* $|\psi\rangle = \sum_{i=1}^{r} \alpha_i |\psi_i^A\rangle |\psi_i^B\rangle$ *be the Schmidt decomposition of* $|\psi\rangle \in \mathcal{A} \otimes \mathcal{B}$. *Then, there exists* $U_{\mathbf{v}}^A \in \mathcal{U}(\mathcal{A})$ *with* $v_k \geq 1$ *(i.e.* $U_{\mathbf{v}}^A$ *has an eigenvalue of multiplicity* k*),* $v_{k'>k} = 0$ *(all eigenvalues of* $U_{\mathbf{v}}^A$ *have multiplicity at most* k*), and* $D(|\psi\rangle\langle\psi|, U_{\mathbf{v}}^A) = 0$ *if and only if* $k \geq r$.

Proof. Suppose $k \geq r$. Then, by defining $\left\{\Pi_j^A\right\}_{\mathbf{v}}^k$ such that $v_k \geq 1$ and $v_{k'>k} = 0$, one can choose a $\left\{\Pi_j^A\right\}_{\mathbf{v}}^k$ such that Equation (6.40) holds for $\rho = |\psi\rangle\langle\psi|$ (i.e. simply project onto $\text{span}(\{|\psi_i^A\rangle\})$). By Theorem 6.11, this implies there exists a $U_{\mathbf{v}}^A$ with $v_k \geq 1$ and $v_{k'>k} = 0$ achieving $D(|\psi\rangle\langle\psi|, U_{\mathbf{v}}^A) = 0$. Conversely, if $k < r$, then clearly no such $\left\{\Pi_j^A\right\}_{\mathbf{v}}^k$ such that Equation (6.40) holds exists. By Theorem 6.11, this implies that no U_A with an eigenvalue of multiplicity at most k and $D(|\psi\rangle\langle\psi|, U_A) = 0$ exists, as desired. □

We close this section with two final comments. First, given Theorem 6.10, one might ask whether a stronger relationship between $D(\rho)$ and $\delta(\rho)$ holds. For example, could it be that $D(\rho) \geq \delta(\rho)$ for all ρ? This simplest type of relationship is ruled out easily via Theorem 6.6 and Equation (6.29), since for $d = 2$ and $p = 2/3$, $D(\rho) = 1/9 \geq \delta(\rho) \approx 0.01614$, while for $d = 50$ and $p = 2/3$, $D(\rho) \approx 0.00627 \leq \delta(\rho) \approx 0.07111$.

Second, note that Theorem 6.11 reduces to Theorem 6.10 if we choose $\mathbf{v} = (M, 0, \ldots, 0)$. This suggests defining a *generalized quantum discord*, denoted $\delta_{\mathbf{v}}(\rho)$, which is analogous to $\delta(\rho)$, except that now we use the class of measurements $\left\{\Pi_j^A\right\}_{\mathbf{v}}$ in the definition of discord (see Equation 5.14). For example, $\delta_{(M,0,\ldots,0)}(\rho) = \delta(\rho)$. We hope the study of $\delta_{\mathbf{v}}(\rho)$ would prove fruitful in its own right.

6.7 Maximally non-classical, yet separable, $(2 \times N)$-dimensional states

In this section, we characterize the set of maximally non-classical, yet separable, $(2 \times N)$-dimensional states of rank at most 2, as quantified by $D(\rho)$. To do so, consider separable state

$$\rho = \sum_{i=1}^{n} p_i |a_i\rangle\langle a_i| \otimes |b_i\rangle\langle b_i|, \tag{6.41}$$

where $\sum_i p_i = 1$, $|a_i\rangle \in \mathbb{C}^2$, $|b_i\rangle \in \mathbb{C}^N$. Via simple algebraic manipulation, one then finds that $D(\rho, U_A)$ for any given $U_A \in \mathcal{U}(\mathcal{A})$ is given by

$$\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j \, |\langle b_i|b_j\rangle|^2 \, (|\langle a_i|a_j\rangle|^2 - |\langle a_i|U_A|a_j\rangle|^2)}. \tag{6.42}$$

We begin by proving a simple but useful upper bound on $D(\rho)$ which depends solely on n.

Lemma 6.13. *Let ρ be a separable state as given by Equation (6.41). Then $D(\rho) \leq 1 - \max_i p_i \leq 1 - \frac{1}{n}$.*

Proof. Assume WLOG that $\max_i p_i = p_1$. Then $1/n \leq p_1 \leq 1$. Choose any $U_A \in \mathcal{U}(\mathcal{A})$ such that $|a_1\rangle$ is an eigenvector of U_A. Then any term in the double sum of Equation (6.42) in which $|a_1\rangle$ appears vanishes. We can hence loosely upper bound the value of Equation (6.42) by $\sqrt{(\sum_{i\neq 1, j\neq 1} p_i p_j)} = 1 - p_1$. Recalling that $p_1 \geq 1/n$ yields the desired bound. □

When $n = 2$, i.e. when ρ is rank at most two, observe from Lemma 6.13 that $D(\rho) \leq 1/2$, and this is attainable only when $p_1 = p_2 = 1/2$. We now show that this bound can indeed be saturated, and characterize all states with $n = 2$ that do so.

Lemma 6.14. *Let ρ be a separable state as in Equation (6.41) with $p_1 = p_2 = 1/2$. Then $D(\rho) = 1/2$ if and only if $|\langle a_1|a_2\rangle| = 1/\sqrt{2}$ and $\langle b_1|b_2\rangle = 0$.*

Proof. Since by Lemma 6.1, $D(\rho)$ is invariant under local unitaries, we can assume without loss of generality that $|a_1\rangle = |0\rangle$, $|b_1\rangle = |0\rangle$, $|a_2\rangle = \cos\frac{\beta}{2}|0\rangle + \sin\frac{\beta}{2}|1\rangle$ and $|b_2\rangle = \sum_{i=0}^{N-1} \alpha_i |i\rangle$ for $\beta \in [0, \pi]$ and $\alpha_i \in \mathbb{R}$ with $\sum_i \alpha_i^2 = 1$, i.e. we can rotate the local states so as to eliminate relative phases. Further, since $U_A \in \mathrm{RU}(\mathcal{A})$ in Equation (6.42), we can write $U_A = 2|u\rangle\langle u| - I$ for some $|u\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle$, where $\theta, \phi \in [0, 2\pi)$. Via the latter, we can rewrite Equation (6.42) as:

$$\frac{1}{2}\sqrt{\sum_{i,j=1}^{2} \langle b_i|b_j\rangle^2 (\langle a_i|a_j\rangle^2 - |\langle a_i|a_j\rangle - 2\langle a_i|u\rangle\langle u|a_j\rangle|^2)}. \tag{6.43}$$

Letting Δ denote the expression under the square root above, we have by substituting in our expressions for $|a_1\rangle$, $|a_2\rangle$, $|b_1\rangle$, $|b_2\rangle$, and $|u\rangle$ and algebraic manipulation that

$$\begin{aligned} \Delta \;=\; & \alpha_0^2 \left[2\cos\beta\sin^2\theta - \sin\beta\sin(2\theta)\cos\phi\right] + \\ & 1 + \sin^2\theta - (\cos\beta\cos\theta + \sin\beta\sin\theta\cos\phi)^2. \end{aligned} \tag{6.44}$$

Our goal is to maximize Δ with respect to α_0 and β (which define ρ), and then minimize with respect to θ and ϕ (which define U_A). Observe now that choosing $\phi = \theta = 0$ reduces Equation (6.44) to $\Delta = 1 - \cos^2\beta$. Hence, unless $\beta = \pi/2$ (i.e. $|\langle a_1|a_2\rangle| = 1/\sqrt{2}$), we can always achieve $D(\rho) < 1/2$. Thus, set $\beta = \pi/2$. Consider next $\phi = 0$, and leave θ unassigned. Then, Equation (6.44) reduces to $\Delta = 1 - \alpha_0^2 \sin(2\theta)$, from which it is clear that unless $\alpha_0 = 0$ (i.e. $\langle b_1|b_2\rangle = 0$), we can always achieve $D(\rho) < 1/2$. Plugging these values of α and β into Equation (6.44), we have $\Delta = 1 + \sin^2\theta \sin^2\phi$, from which the claim follows. □

For two-qubit ρ, we thus have that with respect to $D(\rho)$ and the geometric discord, the maximally non-classical two qubit states of rank at most two are, up to local unitaries,

$$\frac{1}{2}|0\rangle\langle 0| \otimes |0\rangle\langle 0| + \frac{1}{2}|+\rangle\langle +| \otimes |1\rangle\langle 1|,$$

where $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. As mentioned earlier, this matches known results with respect to the relative entropy of quantumness [110]. However, the latter analysis is not as general as it begins by with the assumption that $\langle b_1|b_2\rangle = 0$, whereas we allow arbitrary $|b_1\rangle, |b_2\rangle$. It would be interesting to know whether this analysis can be extended to arbitrary rank two-qubit states.

Acknowledgements for this chapter. We thank Gerardo Adesso, Dagmar Bruß, Davide Girolami and Marco Piani for helpful discussions.

Chapter 7

All non-classical correlations can be activated into distillable entanglement

This chapter is based on [208]:

M. Piani, S. Gharibian, G. Adesso, J. Calsamiglia, P. Horodecki and A. Winter. All non-classical correlations can be activated into distillable entanglement. *Physical Review Letters*, 106:220403, 2011, DOI: 10.1103/PhysRevLett.106.220403,

In this chapter, we introduce a protocol through which general non-classical multipartite correlations can be mapped or "activated" into bipartite entanglement. In particular, we provide an operational interpretation for the measure of non-classicality known as the relative entropy of quantumness, showing that it quantifies the minimum distillable entanglement generated between the initial system and ancillae in our protocol. Moreover, we show the following surprising fact: That mixed entangled states can be arbitrarily more non-classical than separable and pure entangled states.

7.1 Introduction and results

The study of quantum correlations has traditionally focused on entanglement [151]. In particular, it is generally believed that entanglement is a necessary resource for quantum

computers to outperform their classical counterparts. Indeed, it has been shown that for the setting of *pure*-state computation, the amount of entanglement present must grow with the system size for an exponential speed-up to occur [160]. In the context of *mixed*-state quantum information processing, however, there are surprising quantum computational and communication feats which are seemingly impossible to achieve with a classical computer, and yet can be attained with a quantum computer using little or no entanglement. Examples include the DQC1 model of computing [174] and the locking of classical correlations [87]; see Section 1.6.2 for a brief exposition. In the case of locking, for example, the task involved is impossible classically, and yet the quantum states used are *separable*. This raises the question: What is the fundamental resource enabling such feats?

One plausible explanation is the presence in (generic [94]) quantum states of non-classical correlations *beyond* entanglement. Indeed, as outlined in Section 1.6.2, much attention has recently been devoted to understanding and quantifying such correlations for this reason [203, 138, 187, 119, 196, 118, 217, 185, 54, 209, 188, 94, 13, 216]. In particular, the separable quantum states of the systems involved in DQC1 and the locking protocol have been shown to possess non-zero amounts of such correlations (see e.g. [80, 78]), as measured by the *quantum discord* [203, 138]. The latter strives to capture non-classical correlations beyond entanglement and has recently received operational interpretations in terms of the quantum state merging protocol [62, 190], but is unfortunately not a *faithful* measure (here, a *faithful* measure achieves a non-zero value of zero if and only if a state is "non-classical"). A more accurate quantification of non-classical correlations is provided by the so-called *relative entropy of quantumness* (REQ) [54, 187, 118, 217, 196], defined as the minimum distance, in terms of relative entropy, between a multipartite quantum state and the closest strictly classically correlated state (see Definition 7.1). Such a measure is faithful [118], symmetric under permutation of the subsystems, and enables a unified approach to the quantification of classical, separable and entangled correlations [196]. However, to date it still lacks an operational interpretation.

More generally, in this chapter, we ask the following question: *Is there a protocol by which general non-classical correlations produce a physically relevant effect that distinguishes them from purely classical ones?*

It turns out that the answer to the above question is not only *yes*, but that among other results, the protocol we derive lends the desired operational interpretation to the REQ.

Our results: In order to summarize our results, recall first from Equation (1.128) the definition of a *strictly classically correlated* or *classical* state in the bipartite setting. For completeness, we state the generalization of this definition to the multipartite setting

below [209].

Definition 7.1 (Strictly classically correlated quantum state)**.** *Let $\rho \in \mathcal{D}((\mathbb{C}^d)^{\otimes n})$, i.e. ρ acts on n d-dimensional systems. Let $\mathcal{B}_i := \{|b_i(j)\rangle\}_{j=0}^{d-1}$ denote some orthonormal basis for $\mathbb{C}^d$ for the ith system, and let $\mathcal{B}$ denote the orthonormal basis*

$$\{|b(k)\rangle := |b_1(k_1)\rangle|b_2(k_2)\rangle \cdots |b_n(k_n)\rangle\} \tag{7.1}$$

for the entire space $(\mathbb{C}^d)^{\otimes n}$ formed by taking tensor products of all elements in bases $\{\mathcal{B}_i\}_{i=1}^n$. Here, $k := k_1 k_2 \cdots k_n$ is a number written in base d. We henceforth use the notation $\mathcal{B}$ to refer to such a local product basis*. Then, an n-qudit state ρ is* strictly classically correlated*, or* classical*, if there exists a local product basis $\mathcal{B}$ with respect to which ρ is diagonal.*

Recall that classical states correspond to the embedding of a multipartite classical probability distribution into the quantum formalism, and that states not of the form above are called *non-classical*. We now summarize our results as follows.

1. An "activation" protocol for non-classical correlations. Our first result is a protocol through which non-classical correlations are mapped into entanglement. Roughly, given an input state $\rho \in \mathcal{D}((\mathbb{C}^d)^{\otimes n})$, the protocol first introduces an ancilla state $|0 \cdots 0\rangle \in (\mathbb{C}^d)^{\otimes n}$. We then show that ρ is non-classically correlated if and only if applying local CNOT gates with system i of ρ as control and system i of the ancilla as target always creates (distillable) entanglement across the system-ancilla split, *even if* one adversarially applies local changes of basis to ρ before applying the CNOT gates (Theorem 7.3).

We thus not only have a physical effect arising from non-classical correlations, as desired, but also an entire framework for designing non-classicality measures. Specifically, for each choice of entanglement measure one applies across the system-ancilla gap after the protocol is run, we have the potential for a new non-classicality measure for system ρ.

2. Connections to non-classicality measures. As mentioned above, by applying our favorite entanglement measure across the system-ancilla cut after our protocol is run, we have the potential for discovering new non-classicality measures for the initial system ρ. In this vein, we first find that applying the entanglement measure *distillable entanglement* [210], we obtain a non-classicality measure we call the *minimum distillable entanglement potential*, which turns out to equal the REQ (Corollary 7.5). We thus have an operational interpretation for the REQ. We also consider the *negativity* [241] as an entanglement measure, obtaining various results of interest here (Section 7.4.2).

3. Mixedness versus entanglement in non-classicality. Our final result studies the minimum distillable entanglement potential (or equivalently, REQ). As might be expected, we first find that according to this non-classicality quantifier, pure entangled states are strictly "more non-classical" than separable states. However, perhaps surprisingly, we next show that in the asymptotic setting, (1) separable states can be as non-classical as pure entangled states (Theorem 7.10), and (2) *mixed* entangled states can be much more non-classical than pure entangled states (Theorem 7.11)! This suggests that non-classical correlations arise not just from the superposition principle of quantum mechanics, as is the case with (pure state) entanglement, but also due to the non-commutative nature of quantum physics. Our proofs here use ideas similar to known concentration of measure arguments [136, 137].

Previous work. We refer the reader to Section 1.6.2 for a brief introduction to non-classical correlations. With regards to this chapter, we remark that after completion of the paper this chapter is based on, we became aware of related results by Streltsov, Kampermann and Bruß [231]. They show that the quantumness of correlations (as measured, for example, by the quantum discord) is also related to the minimum entanglement generated between system and apparatus in a partial measurement process. In light of those results, our findings can be understood also as dealing with the interplay between system-apparatus entanglement and non-classicality of correlations when realizing local measurements.

Discussion and open questions. The study of general non-classical correlations is currently a burgeoning area, but in many ways such correlations are still not well-understood. Our activation protocol lends new insight into the nature of these correlations by furnishing them with a new operational meaning in terms of resources for *entanglement generation*. One natural and interesting open question is whether the ideas behind the protocol could lead to novel applications in quantum computation and information.

Furthermore, our novel framework for non-classicality measures reduces the problem of non-classicality quantification to the more familiar setting of entanglement quantification, for which a multitude of tools for analysis are already known (see e.g. [151]). An open question here is what further known non-classicality quantification schemes can be obtained as arising through our framework?

Finally, that mixing can actually help *surpass* the quantumness of pure-state entanglement, and that the latter can be asymptotically matched by fully separable states is, in our opinion, quite a surprising result. It would be good to better understand the non-commutative nature of states in a quantum mixture, both from the perspective of non-

classical correlations, as well as with regard to computational and information theoretic feats.

Organization of chapter. In Section 7.3, we describe our activation protocol, and show how it yields a connection between entanglement and non-classical correlations. Section 7.4 then exploits this connection further by introducing an entire family of non-classicality quantifiers, demonstrating along the way an operational interpretation of the REQ. In Section 7.5, we show two surprising results in systems of large local dimensions: That mixed separable states can be asymptotically as non-classical as pure maximally entangled states, and that mixed entangled states can be asymptotically twice as non-classical as pure maximally entangled states.

7.2 Preliminaries

We now state notation and a lemma specific to this chapter. Regarding notation, given a local product basis $\mathcal{B} = \{|b(k)\rangle\}$ and multipartite quantum state ρ, we define

$$\rho^{\mathcal{B}} := \sum_k |b(k)\rangle\langle b(k)|\rho|b(k)\rangle\langle b(k)|, \tag{7.2}$$

and

$$\rho^{\mathcal{B}}_{kl} := \langle b(k)|\rho|b(l)\rangle. \tag{7.3}$$

We next state Levy's Lemma, which is useful in Section 7.5. For this, we first define the *Lipschitz constant* of a function f. Given function $f : X \mapsto Y$ for metric spaces (X, d_X) and (Y, d_Y), where d_X and d_Y are metrics on the sets X and Y, respectively, we say that f has Lipschitz constant $m \geq 0$ if the distance between any two input points in X does not increase by more than m after going through f. In other words, for all $x_1, x_2 \in X$,

$$d_Y(f(x_1), f(x_2)) \leq m \cdot d_X(x_1, x_2). \tag{7.4}$$

Then, for $\mathbb{S}^k$ the k-sphere and $\mathbb{E}(f)$ the expected value of function f, we can state the following useful Lemma, known as Levy's Lemma.

Lemma 7.2 (Levy's Lemma, see e.g. [136])**.** *Let $f : \mathbb{S}^k \mapsto \mathbb{R}$ be a function whose Lipschitz constant with respect to the Euclidean norm is $m \geq 0$. Let $x \in \mathbb{S}^k$ be chosen uniformly at random. Then, for some constant $c > 0$,*

$$\Pr\left(f(x) - \mathbb{E}(f) \gtrless \pm\alpha\right) \leq 2\exp\left(\frac{-c(k+1)\alpha^2}{m^2}\right). \tag{7.5}$$

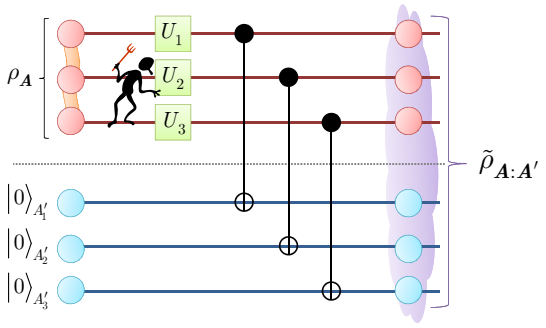


Figure 7.1: (Color online) Scheme of the activation protocol for $n = 3$.

7.3 The activation protocol

We now describe our protocol for the activation of non-classical correlations, which maps relatively "not-well-understood" non-classical correlations into "more familiar" bipartite entanglement, allowing one to employ tools from entanglement theory [151] to study general non-classical correlations. The protocol can be thought of as a game between an adversary and n players, where the n players together aim to generate an entangled state between a system A they control and an ancillary system A', and the adversary's goal is to thwart their efforts by locally rotating each subsystem of A before system and ancilla undergo a pre-defined interaction.

More precisely, the protocol proceeds as follows (see Figure 7.1). We consider n players $\mathcal{P}_i$, each controlling a system-ancilla pair of qudits (A_i, A'_i). We indicate by A the joint register $A_1, \ldots, A_n$, henceforth called the "system", and by A' the joint register $A'_1, \ldots, A'_n$, henceforth called the "ancilla". The initial state of the total $2n$ qudits is a tensor product $\rho_{AA'} = \rho_A \otimes |0\rangle\langle 0|^{\otimes n}_{A'}$. For a given ρ_A, an adversary is first allowed to apply a local unitary U_i of his choice to each A_i. With the adversary's turn complete, each player $\mathcal{P}_i$ now lets their subsystem A_i (control qudit) interact with the corresponding ancillary party A'_i (target qudit) via a CNOT gate, whose action on the computational basis states $|j\rangle|j'\rangle$ of $\mathbb{C}^d \otimes \mathbb{C}^d$ is defined as $|j\rangle|j'\rangle \mapsto |j\rangle|j' \oplus j\rangle$, with $\oplus$ denoting addition modulo d. The final state of system plus ancilla is

$$\rho^f_{AA'} = V(\rho_A \otimes |0\rangle\langle 0|^{\otimes n}_{A'})V^\dagger\,, \tag{7.6}$$

where $V = CNOT_{AA'}(U_A \otimes I_{A'})$, $U_A = \otimes_{i=1}^n U_i$ and $CNOT_{AA'} = \bigotimes_{i=1}^n CNOT_{A_iA'_i}$. We

ask: At the end of the protocol, have the n players succeeded in generating bipartite entanglement across the split $A : A'$, and, if so, how much entanglement was created? It is natural to expect that the answer will depend on the initial state ρ_A of the n-qudit system. For simplicity of notation, in the remainder of this chapter, we shall take ρ and ρ^f to denote the states ρ_A and $\rho^f_{AA'}$, respectively.

Although we cast the activation protocol as a game, from a more physical perspective our aim is to understand precisely how the nature and amount of correlations between the parts A_i of the system A affects the entanglement that can be created with an ancilla A' via the paradigmatic entangling operation — the CNOT; we are considering here the worst case scenario with respect to the choice of the control bases. We then find the following.

Theorem 7.3. *A state ρ of an n-qudit system is classical if and only if there exists an adversarial choice of local unitaries U_A such that the state ρ^f output by the activation protocol is separable across the system-ancilla (i.e. $A : A'$) split.*

Proof. The "if" part is trivial, as given a strictly classically correlated state, one can choose U_A to rotate the diagonalizing local product basis for $\rho = \sum_i p_i |b(i)\rangle\langle b(i)|$ into the computational basis, so that applying the CNOTs in our protocol straightforwardly yields the separable state

$$\rho^f = \sum_i p_i |i\rangle\langle i|_A \otimes |i\rangle\langle i|_{A'}. \tag{7.7}$$

As for the "only if" part, consider the separable decomposition

$$\rho^f = \sum_i q_i |\psi_i\rangle\langle\psi_i|_A \otimes |\phi_i\rangle\langle\phi_i|_{A'}, \tag{7.8}$$

which exists by hypothesis for some choice of U_A. Since the transformation V in Equation (7.6) is unitary, we must be able to write $\rho = \sum_i q_i |v_i\rangle\langle v_i|_A$ for some ensemble (not necessarily a spectral decomposition) $\{q_i, |v_i\rangle\}$ such that

$$V|v_i\rangle_A|0\rangle_{A'} = |\psi_i\rangle_A|\phi_i\rangle_{A'}. \tag{7.9}$$

Letting $\{|j\rangle\}$ denote the computational basis, we now expand $|v_i\rangle$ in the basis $\left\{U_A^\dagger|j\rangle\right\}$, such that

$$|v_i\rangle_A = \sum_j \alpha_{ij} U_A^\dagger |j\rangle_A, \tag{7.10}$$

from which it follows that

$$V|v_i\rangle_A|0\rangle_{A'} = \sum_j \alpha_{ij} |j\rangle_A|j\rangle_{A'}. \tag{7.11}$$

Combining this with Equation (7.9), we conclude that for all i, there must exist a j such that $\alpha_{ij} = 1$. Denote this value of j as j_i, and note hence that

$$|v_i\rangle_A|0\rangle_{A'} = V^\dagger\left(|j_i\rangle_A|j_i\rangle_{A'}\right) = U_A^\dagger|j_i\rangle_A|0\rangle_{A'}. \tag{7.12}$$

We can now write

$$\rho = \sum_i q_i|v_i\rangle\langle v_i|_A = \sum_i q_i U_A^\dagger|j_i\rangle\langle j_i|_A U_A, \tag{7.13}$$

which is a spectral decomposition for ρ with respect to the computational basis up to local unitary U_A, as desired. □

In other words, the system *always* becomes (for any choice of U_A) entangled with the ancilla as a result of the activation protocol, if and only if the input state of the system is non-classically correlated. This establishes a qualitative *equivalence* between multipartite non-classical correlations among components of a quantum system, and bipartite entanglement between the system and an ancilla.

7.4 Quantifying non-classicality

We now exploit the spirit of Theorem 7.3 further to *quantify*, rather than simply detect, the presence of non-classical correlations in a quantum state. To do so, our approach is to apply entanglement measures across the $A : A'$ split to study the amount of entanglement generated whenever A is initially in a non-classically correlated state. It is worth remarking here that this framework is general enough to possibly uncover a full zoology of non-classicality measures, as each choice of a different entanglement monotone [210] we adopt (at the output) has the potential to lead to a unique non-classicality measure (for the input state), the association being provided exactly by the activation protocol.

More precisely, let E denote some entanglement measure of choice and ρ^f the system-ancilla state at the end of the protocol as in Equation (7.6), and define by

$$Q_E(\rho) := \min_{U_A} E_{A:A'}(\rho^f) \tag{7.14}$$

the minimum entanglement generated across the $A : A'$ split over all choices of adversarial local unitaries U_A. We call $Q_E(\rho)$ the *minimum entanglement potential* of ρ with respect to E. As a consequence of Theorem 7.3, Q_E is a measure of non-classical correlations for arbitrary multipartite qudit states ρ, induced by the entanglement monotone E. In

fact, the condition $Q_E(\rho) = 0$ perfectly characterizes the set of classically correlated states ρ if E is a *faithful* entanglement measure (i.e. if E vanishes only for separable states). However, even certain non-faithful entanglement measures can be plugged in to obtain a faithful measure of non-classical correlations. The reason is that the output state ρ^f has the so-called *maximally correlated* form [212] between A and A'; namely,

$$\rho^f = \sum_{kl} \rho^{\mathcal{B}}_{kl} |k\rangle\langle l|_A \otimes |k\rangle\langle l|_{A'} \tag{7.15}$$

with $\rho^{\mathcal{B}}_{kl} = \langle b(k)|\rho|b(l)\rangle$, $|b(k)\rangle = U_A^\dagger |k\rangle$ and $|k\rangle = |k_1\rangle|k_2\rangle\cdots|k_n\rangle$. We now exploit this observation in the next section.

7.4.1 Minimum distillable entanglement potential

Let us consider the non-faithful but physically motivated distillable entanglement E_{D} [210] as a bipartite entanglement monotone (recall E_{D} is non-faithful as it vanishes on so-called bound entangled states). Note that the precise definition of E_{D} is not required here; rather we utilize results of [140] linking $E_{\mathrm{D}}(\rho)$ to the relative entropy of entanglement. Specifically, we have the following.

Theorem 7.4. *The minimum distillable entanglement potential $Q_{E_D}(\rho)$ equals*

$$Q_{E_{\mathrm{D}}}(\rho) = \min_{\mathcal{B}} \Big(S(\rho^{\mathcal{B}}) - S(\rho) \Big), \tag{7.16}$$

where the minimization is over the choice of local product bases $\mathcal{B}$.

Proof. The claim follows by observing that for any choice of $\mathcal{B}$, the $A : A'$ distillable entanglement of ρ^f is equal to

$$E_{\mathrm{D}}(\rho^f) = S(\mathrm{Tr}_{A'}(\rho^f)) - S(\rho^f) = S(\rho^{\mathcal{B}}) - S(\rho), \tag{7.17}$$

where $S(\sigma) = -\mathrm{Tr}(\sigma \log \sigma)$ is the von Neumann entropy of a state σ. In the first equality we used the results of [140] about distillable entanglement for maximally correlated states — for which it happens to coincide with the relative entropy of entanglement [239, 238]. The second equality is justified by the fact that $\rho^{\mathcal{B}}$ is the state resulting from local projective measurements in the local bases $\mathcal{B}$ on ρ and is unitarily equivalent to $\mathrm{Tr}_{A'}(\rho^f)$ (seen by considering Equation (7.15)), while ρ^f is obtained from ρ via the activation protocol isometry, Equation (7.6). □

This yields the following nice corollary regarding the REQ, which is defined as (see also Section 1.6.2)

$$Q(\rho) = \min_{\text{classical } \sigma} S(\rho\|\sigma), \tag{7.18}$$

for $S(\rho\|\sigma) = \text{Tr}(\rho \log \rho - \rho \log \sigma)$ the relative entropy and where the minimization is over all strictly classically correlated states σ.

Corollary 7.5. *The REQ of ρ equals its minimum distillable entanglement potential, i.e.*

$$Q(\rho) = Q_{E_{\text{D}}}(\rho). \tag{7.19}$$

Proof. The claim follows immediately from the result that, as proven for example in Theorem 2 of [196], the REQ can alternatively be expressed as the expression in Equation (7.16). □

This finding immediately provides a clear-cut *operational interpretation* for the REQ, which therefore emerges as a natural, mathematically sound and physically motivated measure of non-classical correlations in quantum states of arbitrary-dimensional composite systems. The degree of non-classical correlations as quantified by the REQ, a measure whose original definition was purely geometric [196], is quantitatively reinterpreted as the resource power of such correlations for the task of generating distillable entanglement with an ancilla in the worst case scenario. Incidentally, since the REQ is faithful [118], this can be considered an alternate proof of Theorem 7.3.

Before closing this section, we prove a strict upper bound on the non-classicality of *separable* bipartite quantum states with respect to $Q_{E_{\text{D}}}$.

Theorem 7.6. *Consider bipartite separable state $\rho_{AB} = \sum_i p_i |\alpha_i\rangle\langle\alpha_i| \otimes |\beta_i\rangle\langle\beta_i|$, for $\{p_i\}$ a probability distribution and $\{|\alpha_i\rangle\}, \{|\beta_i\rangle\} \subseteq \mathbb{C}^d$. Then, $Q(\rho_{AB}) < \log d$.*

Proof. We have

$$\begin{aligned}
Q(\rho_{AB}) &= \min_{\mathcal{B}} \Big(S(\rho_{AB}^{\mathcal{B}}) - S(\rho_{AB}) \Big) && (7.20)\\
&\leq \min_{\mathcal{B}} \Big(S(\rho_{AB}^{\mathcal{B}}) - S(\rho_A) \Big) && (7.21)\\
&= \min_{\mathcal{B}} \Big(S(\rho_A^{\mathcal{B}_A}) + \sum_i \langle b_A(i)|\rho_A|b_A(i)\rangle S(\sigma_i^{\mathcal{B}_B}) - S(\rho_A) \Big) && (7.22)\\
&\leq \min_{\mathcal{B}_B} \sum_i p_i^A S(\sigma_i^{\mathcal{B}_B}), && (7.23)
\end{aligned}$$

where

$$\sigma_i^{\mathcal{B}_B} := \sum_j \frac{\langle b_A(i)b_B(j)|\rho_{AB}|b_A(i)b_B(j)\rangle}{\langle b_A(i)|\rho_A|b_A(i)\rangle}|b_B(j)\rangle\langle b_B(j)|, \tag{7.24}$$

where $\{p_i^A\}$ are the eigenvalues of ρ_A, the first inequality follows since for any separable state, $S(\rho_{AB}) \geq \max\{S(\rho_A), S(\rho_B)\}$ [201], and the second inequality by choosing $\mathcal{B}_A$ as an eigenbasis of ρ_A (yielding $S(\rho_A^{\mathcal{B}_A}) = S(\rho_A)$).

Suppose now, for sake of contradiction, that this upper bound is equal to $\log d$. Then, it must be the case that $\sigma_i^{\mathcal{B}_B}$ is maximally mixed for all i, implying that ρ_B is also maximally mixed. Reversing the role of A and B, an analogous argument yields that ρ_A must be maximally mixed as well. This means that the basis chosen in the second inequality is arbitrary, and we find that for the last line to be equal to $\log d$, it must be that $\langle b_A(i)b_B(j)|\rho_{AB}|b_A(i)b_B(j)\rangle = 1/d^2$ for all b_A, b_B and all i, j. Thus, $\rho_{AB} = I/d^2$. However, this state is classical, and hence achieves $Q(\rho_{AB}) = 0$, yielding the desired contradiction. □

7.4.2 Negativity of quantumness

The next entanglement monotone we consider in our scheme is the *Negativity* [241]. The latter is defined for a bipartite state ρ_{AB} as $\mathcal{N}(\rho_{AB}) := (\left\| \rho_{AB}^{T_A} \right\|_{\text{tr}} - 1)/2$, for $\rho_{AB}^{T_A}$ the partially transposed state. Plugging $\mathcal{N}$ into our framework, we obtain a non-classicality measure we call the *negativity of quantumness*, $Q_{\mathcal{N}}(\rho)$.

Theorem 7.7. *For the negativity of quantumness, $Q_{\mathcal{N}}$, we have that*

$$Q_{\mathcal{N}}(\rho) = \frac{1}{2}\min_{\mathcal{B}} \sum_{i \neq j} |\rho_{ij}^{\mathcal{B}}|. \tag{7.25}$$

Proof. Thanks to the maximally correlated form of the output of our protocol, by directly applying the definition of the partial transpose to ρ^f, we can calculate the eigenvalues of $(\rho^f)^{T_A}$ as ρ_{ii}^b for all i, and $\pm|\rho_{ij}^b|$ for $i > j$. Thus,

$$\mathcal{N}(\rho^f) = \frac{\left\| (\rho^f)^{T_A} \right\|_{\text{tr}} - 1}{2} = \frac{\sum_{i \neq j} |\rho_{ij}^{\mathcal{B}}|}{2}. \tag{7.26}$$

□

We remark that since, by definition, a non-classical state must have some non-vanishing off-diagonal terms $\rho_{ij}^{\mathcal{B}}$ in any local product basis $\mathcal{B}$, we thus obtain yet another proof of Theorem 7.3, i.e. that ρ^f is entangled for any local rotation U_A if and only if ρ is not classical.

Next, for the special case of *pure* bipartite states $|\psi\rangle$, we find that $Q_{\mathcal{N}}$ has a particularly simple form, in that it reduces to the negativity of $|\psi\rangle$.

Corollary 7.8. *For rank one bipartite states $|\psi\rangle\langle\psi|$,*

$$Q_{\mathcal{N}}(|\psi\rangle\langle\psi|) = \mathcal{N}(|\psi\rangle\langle\psi|). \tag{7.27}$$

Proof. Note first that for $|\psi\rangle = \sum_i \alpha_i |a_i\rangle|b_i\rangle$ the Schmidt decomposition of $|\psi\rangle$, one has $\mathcal{N}(|\psi\rangle\langle\psi|) = \sum_{i\neq j} \alpha_i\alpha_j$. We now show that $Q_{\mathcal{N}}(|\psi\rangle\langle\psi|)$ matches this expression.

For any local product basis $\mathcal{B}$, one can write $|\psi\rangle = \sum_{i,j=0}^{d-1} \alpha_{ij}|b_1(i)\rangle|b_2(j)\rangle$. Then, letting $\rho = |\psi\rangle\langle\psi|$ and beginning from Equation (7.15), straightforwardly applying the definitions of the trace norm and partial transpose yields in Equation (7.26) that

$$\left\| \left(\rho^f\right)^{T_A} \right\|_{\mathrm{tr}} = \left(\sum_{ij} |\alpha_{ij}| \right)^2 . \tag{7.28}$$

Note here that the coefficients α_{ij} are specific to the choice of basis $\mathcal{B}$ — thus, our goal is to choose $\mathcal{B}$ so as to minimize $\sum_{ij} |\alpha_{ij}|$. We claim that this minimizing basis is in fact just the tensor product of the local Schmidt bases for $|\psi\rangle$.

To see this, we use the vec mapping [246], which can be defined such that $\mathrm{vec}(|a\rangle\langle b|) = |a\rangle|b\rangle$ (and analogously, $\mathrm{vec}^{-1}(|a\rangle|b\rangle) = |a\rangle\langle b|$), and the l_1 norm, defined as $\| C \|_{l_1} := \sum_{ij} |C_{ij}|$. Define now

$$C := \mathrm{vec}^{-1}\left(\sum_{ij=0}^{d-1} \alpha_{ij}|b_1(i)\rangle|b_2(j)\rangle \right) = \sum_{ij=0}^{d-1} \alpha_{ij}|b_1(i)\rangle\langle b_2(j)|. \tag{7.29}$$

Then, we have

$$\sum_{ij} |\alpha_{ij}| = \| C \|_{l_1} \geq \| C \|_{\mathrm{tr}} = \sum_i \sigma_i, \tag{7.30}$$

for $\{\sigma_i\}$ the singular values of C. Here, the claim $\| C \|_{l_1} \geq \| C \|_{\mathrm{tr}}$ follows since

$$\begin{aligned} \| C \|_{\mathrm{tr}} &= \max_{0 \preceq M \preceq I} \mathrm{Tr}((2M - I)C) && (7.31) \\ &\leq \max_{0 \preceq M \preceq I} |\langle \mathrm{vec}(2M - I), \mathrm{vec}(C) \rangle| && (7.32) \\ &\leq \max_{0 \preceq M \preceq I} \| \mathrm{vec}(2M - I) \|_{\infty} \| \mathrm{vec}(C) \|_1 && (7.33) \\ &\leq \| C \|_{l_1}, && (7.34) \end{aligned}$$

where the second inequality follows from the Hölder inequality, the third inequality from the fact that $\| \mathrm{vec}(2M - I) \|_{\infty}$ is at most the spectral norm of $2M - I$, and where $\| v \|_{\infty} := \max_i |v_i|$ and $\| v \|_1 := \sum_i |v_i|$.

The final step is to observe that the σ_i are in fact the Schmidt coefficients of $|\psi\rangle$, since if $C = \sum_i \sigma_i |a_i\rangle\langle b_i|$ is the singular value decomposition of C, then $\mathrm{vec}(C) = \sum_i \sigma_i |a_i\rangle|b_i\rangle$ is a Schmidt decomposition for $|\psi\rangle$. Since $\{|a_i\rangle \otimes |b_j\rangle\}$ is a valid local product basis b in which to expand $|\psi\rangle$, by combining Equations 7.28 and 7.30 the claim follows. □

We finally extend our analysis for pure states to the setting of *pseudo-pure* states

$$\rho(\psi, p) := \frac{(1-p)}{d^2} I + p|\psi\rangle\langle\psi| \tag{7.35}$$

where $0 \leq p \leq 1$.

Corollary 7.9. *For pseudo-pure state $\rho(\psi, p)$, we have*

$$Q_{\mathcal{N}}(\rho) = p\mathcal{N}(|\psi\rangle\langle\psi|). \tag{7.36}$$

Proof. Follows immediately from Equation (7.25) and Corollary 7.8 by recalling that $\rho^{\mathcal{B}}_{ij} = \langle b(i)|\rho|b(j)\rangle$. □

Hence, as already observed in, for example, Reference [118], $\rho(\psi, p)$ is non-classical as long as $p > 0$ and ψ is entangled.

7.5 Non-classicality, mixedness, and entanglement

Equipped with a faithful and operational measure of non-classical correlations, $Q_{E_{\mathrm{D}}}$, which we henceforth refer to as Q, we now investigate the interplay between non-classicality,

entanglement and mixedness of general states ρ_A. For the sake of simplicity, from now on we restrict to the bipartite case $A_1 = A$, $A_2 = B$. We begin by setting the stage with a few simple but general observations following from the definition of Q.

For *pure* states $\rho_{AB} = |\psi\rangle\langle\psi|$, $Q(\rho_{AB})$ reduces to the von Neumann entropy of entanglement $S(\rho_A) = S(\rho_B)$ [54], and is thus at most equal to $\log d$. On the other hand, for arbitrary mixed ρ_{AB}, we have that $Q(\rho_{AB})$ is at most $2\log d$, since from Equation (7.18) one has $Q(\rho_{AB}) \leq S(\rho_{AB}\|\rho_A \otimes \rho_B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \equiv I(\rho_{AB})$, where I denotes the mutual information, a measure of *total* correlations. From this and the results of [201], one realizes that for a *separable* state a bound $Q(\rho_{AB}^{\text{sep}}) \leq \log d$ holds. Now recall from Theorem 7.6 that this inequality is always sharp for separable states, i.e. the bound $\log d$ cannot be exactly saturated for separable non-classical states, while it is instead trivially reached by pure maximally entangled states $|\psi\rangle = d^{-1/2}\sum_{j=0}^{d-1}|j\rangle|j\rangle$.

Surprisingly, what we now show is that as $d \to \infty$, this upper bound is in fact *asymptotically* attained by separable states. More precisely, we show that there exist separable states such that $Q(\rho_{AB}^{\text{sep}})/\log d \to 1$ with growing d. Even more intriguingly, we can show that the upper bound on general mixed bipartite states ρ_{AB} is also asymptotically tight; specifically, there exist families of *mixed* states for which as $d \to \infty$, $Q(\rho_{AB})/\log d \to 2$.

More formally, we prove the following two results, where $m := \lceil (\log d)^4 \rceil$.

Theorem 7.10. *Define the random separable state:*

$$\sigma_{AB} = \frac{1}{dm} \sum_{\substack{i=1,\ldots,d \\ j=1,\ldots,m}} |i\rangle\langle i|_A \otimes \left(U_j |i\rangle\langle i| U_j^\dagger\right)_B , \tag{7.37}$$

for unitaries U_j drawn independently from the Haar measure. Then, with high probability, $Q(\sigma_{AB}) \geq \log d - O(\log\log d)$.

Theorem 7.11. *For C a system of dimension m, let $\rho_{AB} = \text{Tr}_C|\psi\rangle\langle\psi|_{ABC}$, where $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathbb{C}^m$ is uniformly distributed (with probability induced by the Haar measure). Then, with high probability, $Q(\rho_{AB}) \geq 2\log d - O(\log\log d)$.*

What these results tell us is that, first, there are separable states that asymptotically (in d) are as non-classical as the most non-classical pure state (which is the maximally entangled state); second, mixed entangled states can be much more non-classical (namely, twice as much) than pure entangled states. We remark that therefore *both* entanglement *and* mixedness are required to "break the barrier" of $\log d$. This goes against the intuition that entanglement by itself is the strongest form of non-classicality: We demonstrate that mixedness also plays a prominent role and can make correlations maximally non-classical.

One possible explanation for these findings may be the following: Traditionally, the study of quantum correlations has focused on quantum entanglement, which arises from the superposition principle of quantum mechanics. Yet, there is another "non-classical" feature of quantum mechanics to be reckoned with; namely, that quantum systems can be in probabilistic mixtures of *non-commuting* states. What Theorem 7.10 thus quantifies is the extent to which non-commutativity alone can give rise to non-classical correlations. When non-commutativity is then combined with the superposition principle, Theorem 7.11 tells us that the non-classical correlations generated are stronger than possible with either principle alone.

We close this section with the proofs of Theorems 7.10 and 7.11.

Proof of Theorem 7.10. The claim will follow by showing that in Equation (7.16), $S(\sigma_{AB}) \leq \log d + \log m$, whereas for d sufficiently large and with high probability, $S\big(\sigma_{AB}^{\mathcal{B}}\big) \geq 2\log d - \text{const.}$ for all $\mathcal{B}$. The first of these bounds is easy to prove — it follows by observing that the rank of σ_{AB} is at most dm.

As for the second bound, note first that

$$S\big(\sigma_{AB}^{\mathcal{B}}\big) = S(\sigma_A^{\mathcal{B}}) + S(\sigma_B^{\mathcal{B}}) - I(\sigma_{AB}^{\mathcal{B}}) = 2\log d - I(\sigma_{AB}^{\mathcal{B}}), \tag{7.38}$$

which follows since $\sigma_A^{\mathcal{B}} = \sigma_B^{\mathcal{B}} = I/d$. Hence, it suffices to show that $I(\sigma_{AB}^{\mathcal{B}}) \leq \text{const.}$. To see this, note that σ_{AB} is almost identical to the information locking states considered in [137] (see Theorem V.1, Equation (64)), which take the form

$$\rho_{AB} = \frac{1}{dm} \sum_{\substack{i=1,\ldots,d \\ j=1,\ldots,m}} |ij\rangle\langle ij|_A \otimes \left(U_j |i\rangle\langle i| U_j^\dagger\right)_B. \tag{7.39}$$

Letting x and y denote random variables corresponding to the outcomes of local measurements X and Y on A and B, respectively, define $I_c(\rho_{AB}) := \max_{X\otimes Y} I(x:y)$. Then, it is known that for large enough d and with high probability over the choice of local unitaries $\{U_j\}$, $I_c(\rho_{AB}) \leq const.$ (specifically, for our choice of m here, set the parameter ϵ in Equation (66) of [137] to scale as $1/\log d$). Observing that σ_{AB} is attainable from ρ_{AB} via a local operation (namely, we trace out the register containing label j in A), and recalling that the mutual information is non-increasing under partial trace completes the proof. □

Proof of Theorem 7.11. The claim will follow by showing that, in Equation (7.16), $S(\rho) \leq \log m$, whereas for d sufficiently large and with high probability, $S\big(\rho^{\mathcal{B}}\big) \geq 2\log d - \text{const.}$ for all $\mathcal{B}$. Again, the first of these bounds follows simply because the rank of ρ is bounded by m.

Now, let M and N denote arbitrary complete von Neumann measurements on A and B, respectively, such that

$$M = \{M_x = |m_x\rangle\langle m_x|\}_{x=1}^{d}, \quad N = \{N_y = |n_y\rangle\langle n_y|\}_{y=1}^{d}. \tag{7.40}$$

For such a measurement M, let $\Pi_M(\sigma)$ denote the completely positive trace-preserving linear map

$$\Pi_M(\sigma) = \sum_{x=1}^{d} M_x \sigma M_x. \tag{7.41}$$

To prove the desired bound of $S\big(\rho^{\mathcal{B}}\big) \geq 2\log d - \text{const.}$, we use the concentration of measure results of [136] (see also [137]). The intuition is as follows. We first consider a *fixed* set of local measurement bases M and N. Then, one can show that for the random state $|\psi\rangle$ and the corresponding state ρ in the statement of the theorem, the expected value of $S(\Pi_M \otimes \Pi_N(\rho))$ is at least roughly $2\log d$. We then convert this into a high-probability statement using Levy's Lemma (Lemma 7.2), which yields that with high probability, $S(\Pi_M \otimes \Pi_N(\rho))$ will indeed be close to its expected value. This was for a *fixed* choice of local measurements M and N — to extend this statement to *all* such measurements, we use the union bound together with a net argument. Specifically, we cast a δ-net over all choices of local measurements M and N, and apply the union bound to conclude that for all measurements from this net, $S(\Pi_M \otimes \Pi_N(\rho))$ will still be close to its expected value with probability bounded away from 1.

To begin, let M and N be a fixed choice of local measurement bases. We first lower bound the expected value of $S(\Pi_M \otimes \Pi_N(\rho))$ over random choices of $|\psi\rangle$ as follows:

$$S\big(\Pi_M \otimes \Pi_N(\rho)\big) \geq S_2\big(\Pi_M \otimes \Pi_N(\rho)\big) = -\log \sum_{x,y=1}^{d} \left[\mathrm{Tr}\left(|\psi\rangle\langle\psi|(M_x \otimes N_y \otimes I)\right)\right]^2, \tag{7.42}$$

where $S_2(\sigma) = -\log(\mathrm{Tr}(\sigma^2))$ is the quantum Renyi entropy of order 2, and the last equality follows since for rank one M_x and N_y,

$$\mathrm{Tr}\left[(M_x \otimes N_y \rho)^2\right] = \left[\mathrm{Tr}(M_x \otimes N_y \rho)\right]^2. \tag{7.43}$$

Hence,

$$\begin{aligned} \mathbb{E}_\psi\left[S\big(\Pi_M \otimes \Pi_N(\rho)\big)\right] &\geq -\log \mathbb{E}_\psi \left[\sum_{x,y=1}^{d} \left[\mathrm{Tr}(|\psi\rangle\langle\psi|(M_x \otimes N_y \otimes I))\right]^2\right] &(7.44)\\ &= -\log\left(d^2 \mathbb{E}_\psi\left[\left[\mathrm{Tr}(|\psi\rangle\langle\psi|(|0\rangle\langle 0| \otimes |0\rangle\langle 0| \otimes I))\right]^2\right]\right) &(7.45) \end{aligned}$$

where the first statement follows by the convexity of $-\log$, and the second since the distribution of $|\psi\rangle$ is invariant under unitaries. Now,

$$
\begin{aligned}
& d^2 \mathbb{E}_\psi \left[\left[\mathrm{Tr}(|\psi\rangle\langle\psi|_{ABC} (|0\rangle\langle 0|_A \otimes |0\rangle\langle 0|_B \otimes I_C)) \right]^2 \right] && (7.46) \\
= \; & d^2 \mathbb{E}_\psi \left[\mathrm{Tr}(|\psi\rangle\langle\psi|_{ABC} \otimes |\psi\rangle\langle\psi|_{A'B'C'} (|00\rangle\langle 00|_{AA'} \otimes |00\rangle\langle 00|_{BB'} \otimes I_{CC'})) \right] && (7.47) \\
= \; & d^2 \mathrm{Tr}(\mathbb{E}_\psi \left[|\psi\rangle\langle\psi|_{ABC} \otimes |\psi\rangle\langle\psi|_{A'B'C'} \right] (|00\rangle\langle 00|_{AA'} \otimes |00\rangle\langle 00|_{BB'} \otimes I_{CC'})) && (7.48) \\
= \; & \frac{1}{m(d^2 m + 1)} \mathrm{Tr}([I_{ABC:A'B'C'} + W_{ABC:A'B'C'}] (|00\rangle\langle 00|_{AA'} \otimes |00\rangle\langle 00|_{BB'} \otimes I_{CC'})) \\
= \; & \frac{m+1}{d^2 m + 1}, && (7.49)
\end{aligned}
$$

where the first equality uses the fact that $(\mathrm{Tr}(AB))^2 = \mathrm{Tr}[(A \otimes A)(B \otimes B)]$, and the third equality follows since $\mathbb{E}_\psi \left[|\psi\rangle\langle\psi| \otimes |\psi\rangle\langle\psi| \right]$ is proportional to $I + W$ for $W = \sum_{ij} |i\rangle\langle j| \otimes |j\rangle\langle i|$ the swap gate. (One way to see the latter is to note that $\sigma = \mathbb{E}_\psi \left[|\psi\rangle\langle\psi| \otimes |\psi\rangle\langle\psi| \right]$ is invariant under $U \otimes U$ for any unitary U, and is hence a Werner state [254]. Werner states, in turn, can be written as mixtures of the projectors onto the symmetric and antisymmetric spaces, $\Pi_S = (I + W)/2$ and $\Pi_A = (I - W)/2$, respectively. Observing that $\mathrm{Tr}(\sigma \Pi_A) = 0$ yields the claim.) Substituting Equation (7.49) in Equation (7.45) thus yields

$$
\mathbb{E}_\psi \left[S\big(\Pi_M \otimes \Pi_N(\rho)\big) \right] \geq \log \frac{d^2 m + 1}{m+1} \geq 2 \log d - \log \left(1 + \frac{1}{m} \right) =: \Delta. \tag{7.50}
$$

With a lower bound on the expected value of $S(\Pi_M \otimes \Pi_N(\rho))$ in hand, we would now like to show that with high probability, $S(\Pi_M \otimes \Pi_N(\rho))$ indeed takes a value close to its expected value. To show this, we apply Levy's Lemma (Lemma 7.2), which requires an upper bound on the Lipschitz constant of the entropy $S\big(\Pi_M \otimes \Pi_N(\rho)\big)$. The latter is given by the proof of Lemma III.2 of [136], which demonstrates an upper bound on the constant of $\sqrt{8} \log d$. Thus, by Levy's Lemma:

$$
\Pr \left\{ S\big(\Pi_M \otimes \Pi_N(\rho)\big) < \Delta - \epsilon \right\} \leq \exp \left(- \frac{c \epsilon^2 d^2 m}{(\log d)^2} \right), \tag{7.51}
$$

for some constant $c > 0$. This shows that the entropy is indeed large with high probability for a fixed choice of local measurement basis $M \otimes N$.

To extend this to *all* local measurement bases M and N, suppose we had a net of T basis pairs M^t and N^t for $t \in [T]$ able to approximate the quantity $S\big(\Pi_M \otimes \Pi_N(\rho)\big)$ for *any* local measurement $M \otimes N$ within precision 2ϵ. Then, by Equation (7.51) and the union

bound, we would have:

$$\Pr\left\{\exists t \text{ s.t. } S\big(\Pi_{M^t} \otimes \Pi_{N^t}(\rho)\big) < \Delta - \epsilon\right\} \leq T \exp\left(-\frac{c\epsilon^2 d^2 m}{(\log d)^2}\right), \tag{7.52}$$

implying

$$\Pr\left\{\exists M \otimes N \text{ s.t. } S\big(\Pi_M \otimes \Pi_N(\rho)\big) < \Delta - 3\epsilon\right\} \leq T \exp\left(-\frac{c\epsilon^2 d^2 m}{(\log d)^2}\right). \tag{7.53}$$

Thus, if such a 2ϵ-net with small enough T exists, then we are done. Indeed, Lemma 7.12 shows that such a 2ϵ-net exists with

$$T \leq \left(\frac{c' d^{3/2} (\log d)^2}{\epsilon^2}\right)^{4d^2} \tag{7.54}$$

for $c' \in \Theta(1)$. For this value of T, since we set $m = \lceil (\log d)^4 \rceil$, the probability on the right side of Equation (7.53) is bounded away from 1 for large enough d, completing the proof. □

In order to complete the proof of Theorem 7.11, we finally show three lemmas required for the net argument above.

Lemma 7.12. *For any constant $\epsilon > 0$, there exists a set $S := \{M^t \otimes N^t\}_{t=1}^T$ of T local measurement bases (where M^t and N^t are rank one von Neumann measurements each acting on d-dimensional spaces) with*

$$T \leq \left(\frac{c d^{3/2} (\log d)^2}{\epsilon^2}\right)^{4d^2}, \tag{7.55}$$

(for $c > 0$ a constant) such that for all $\rho \in \mathcal{D}(\mathbb{C}^d \otimes \mathbb{C}^d)$ and local measurements $M \otimes N$, there exists a $M^t \otimes N^t \in S$ such that

$$\left|S\big(\Pi_M \otimes \Pi_N(\rho)\big) - S\big(\Pi_{M^t} \otimes \Pi_{N^t}(\rho)\big)\right| \leq 2\epsilon. \tag{7.56}$$

Proof. To construct the set S, we embed each local measurement basis into a unitary matrix, and then cast a net over unitary matrices. Specifically, recall that each local measurement is described by an orthonormal basis $M = \{|b_i\rangle\}_{i=1}^d$. Then, by arranging the vectors $|b_i\rangle$ as columns of a matrix, we obtain a $d \times d$ unitary matrix, denoted U_M, which

rotates the standard basis to $\{|b_i\rangle\}_{i=1}^d$. By Lemma 7.13, there exists a δ-net (with respect to the spectral norm) for $\mathcal{U}(\mathbb{C}^d)$ of size

$$T_0 \leq \left(\frac{c' d^{3/2}}{\delta}\right)^{2d^2}. \tag{7.57}$$

Thus, by setting $\delta = \epsilon^2/32(\log d)^2$ and picking T_0 elements $\{M^s\}$ and T_0 elements $\{N^t\}$, we obtain our set S of local measurement bases with size $T = T_0^2$, as in the statement of our claim.

We now show that S is a 2ϵ-net. Let M and N be arbitrary local measurements with corresponding unitaries U_M and U_N. Then, there exist U_{M^s} and U_{N^t} in the net from Lemma 7.13 such that $\| U_M - U_{M^s} \|_\infty \leq \delta$ and $\| U_N - U_{N^t} \|_\infty \leq \delta$. Let V denote the unitary mapping basis $M^s \otimes N^t$ to basis $M \otimes N$.

Now, if it were true that for all $\rho \in \mathcal{D}(\mathbb{C}^d \otimes \mathbb{C}^d)$,

$$\left\| V\rho V^\dagger - \rho \right\|_{\mathrm{tr}} \leq 4\delta, \tag{7.58}$$

then our desired 2ϵ-net property would follow since

$$\begin{aligned}
\left|S\big(\Pi_M \otimes \Pi_N(\rho)\big) - S\big(\Pi_{M^t} \otimes \Pi_{N^t}(\rho)\big)\right| &= \left|S\big(\Pi_M \otimes \Pi_N(\rho)\big) - S\big(\Pi_M \otimes \Pi_N(V\rho V^\dagger)\big)\right| \\
&\leq H(2\delta, 1-2\delta) + 2\delta \log d^2 && (7.59)\\
&\leq (2\sqrt{2\delta} + 2\delta)\log d^2 && (7.60)\\
&= \epsilon + \frac{\epsilon^2}{8\log d} && (7.61)\\
&\leq 2\epsilon && (7.62)
\end{aligned}$$

for large enough d (or alternatively for $0 < \epsilon < 1$). Here, the second inequality follows from the estimate for the Shannon entropy $H(x, 1-x) \leq 2\sqrt{x(1-x)}$, and the first inequality uses the Fannes-Audenaert inequality [31], which states that for $\rho, \sigma \in \mathcal{D}(\mathbb{C}^d)$ with $r := \| \rho - \sigma \|_{\mathrm{tr}}/2$,

$$|S(\rho) - S(\sigma)| \leq r \log(d-1) + H(r, 1-r). \tag{7.63}$$

Thus, it remains to show that Equation (7.58) indeed holds. To see this, note first that by Lemma 4.4,

$$\| U_M \otimes U_N - U_{M^s} \otimes U_{N^t} \|_\infty \leq \| U_M - U_{M^s} \|_\infty + \| U_N - U_{N^t} \|_\infty \leq 2\delta, \tag{7.64}$$

where note the columns of $U_M \otimes U_N$ are now elements of the local product basis $M \otimes N$. We thus have

$$2\delta \geq \| U_M \otimes U_N - U_{M^s} \otimes U_{N^t} \|_\infty = \left\| U_M \otimes U_N - V^\dagger (U_M \otimes U_N) \right\|_\infty = \| I - V \|_\infty , \tag{7.65}$$

where the last equality follows since the spectral norm is invariant under unitaries. But this implies

$$\begin{aligned} \left\| V\rho V^\dagger - \rho \right\|_{\mathrm{tr}} &= \left\| V\rho V^\dagger - V\rho + V\rho - \rho \right\|_{\mathrm{tr}} && (7.66)\\ &\leq \left\| \rho(V^\dagger - I) \right\|_{\mathrm{tr}} + \| (V - I)\rho \|_{\mathrm{tr}} && (7.67)\\ &\leq \| \rho \|_{\mathrm{tr}} \left\| V^\dagger - I \right\|_\infty + \| V - I \|_\infty \| \rho \|_{\mathrm{tr}} && (7.68)\\ &\leq 4\delta, && (7.69) \end{aligned}$$

where the first inequality follows from the triangle inequality, the second from the fact that for Schatten p-norms, $\| ABC \|_p \leq \| A \|_\infty \| B \|_p \| C \|_\infty$ (see Section 1.3), and the third inequality from Equation 7.65. This concludes the proof.

□

Lemma 7.13. *For any constant $\delta > 0$, there exists a set $S := \{U_t\}_{t=1}^{T_0}$ of unitaries $U_t \in \mathcal{U}(\mathbb{C}^d)$ with*

$$T_0 \leq \left(\frac{cd^{3/2}}{\delta} \right)^{2d^2}, \tag{7.70}$$

(for $c > 0$ a constant) and such that for all $U \in \mathcal{U}(\mathbb{C}^d)$, there exists a $U_t \in S$ satisfying $\| U - U_t \|_\infty \leq \delta$.

Proof. For any unitary $U \in \mathcal{U}(\mathbb{C}^d)$, the idea is to replace the columns of U with vectors taken from a net on the set of pure states in $\mathbb{C}^d$. Of course, the resulting operator U' is in general not unitary — however, this can be corrected by an appropriate orthogonalization procedure inspired by the "pretty good measurement" [135], yielding a unitary U'' such that $\| U - U'' \|_\infty \leq \delta$, as desired.

More specifically, by Lemma III.6 of [136], there exists an ϵ-net N on the set of pure state vectors in $\mathbb{C}^d$ (with respect to the Euclidean norm) such that $|N| \leq (5/\epsilon)^{2d}$. Set $\epsilon = \frac{\delta}{6d^{3/2}}$. Now, for each column $|b_i\rangle$ of a given unitary U, we first find an ϵ-close vector $|b_i'\rangle \in N$, and embed the latter as columns into a matrix U'. Of course, the vectors $\{|b_i'\rangle\}$ are not orthogonal in general, so U' is not unitary. To correct this, define the operator $B = \sum_{i=1}^d |b_i'\rangle\langle b_i'|$ and let

$$|b_i''\rangle = B^{-1/2} |b_i'\rangle. \tag{7.71}$$

Note that if the $|b_i'\rangle$ are linearly independent, then B is invertible, and moreover $\{|b_i''\rangle\}_{i=1}^d$ is an orthonormal basis since

$$\sum_{i=1}^{d} |b_i''\rangle\langle b_i''| = \sum_{i=1}^{d} B^{-1/2}|b_i'\rangle\langle b_i'|B^{-1/2} = B^{-1/2}BB^{-1/2} = I. \tag{7.72}$$

Note that by Lemma 7.14 below, the $|b_i'\rangle$ are indeed linearly independent (for large enough d) for our choice of $\epsilon \in \Theta(d^{-3/2})$.

Now, in order to show that this construction constitutes a δ-net, we must show that $\| U - U'' \|_\infty \leq \delta$. To do so, we first bound $\| |b\rangle - |b''\rangle \|_2$ as

$$\| |b\rangle - |b''\rangle \|_2 \leq \| |b\rangle - |b'\rangle \|_2 + \| |b'\rangle - |b''\rangle \|_2 \leq \epsilon + \left\| (I - B^{-\frac{1}{2}})|b'\rangle \right\|_2 \leq \epsilon + \left\| I - B^{-\frac{1}{2}} \right\|_\infty, \tag{7.73}$$

where the second inequality follows from our ϵ-net, and the third inequality from the definition of the spectral norm. To bound this latter quantity, note that since

$$\| |b_i'\rangle\langle b_i'| - |b_i\rangle\langle b_i| \|_\infty \leq \| |b_i'\rangle\langle b_i'| - |b_i\rangle\langle b_i| \|_{\mathrm{tr}} \leq 2 \| |b_i'\rangle - |b_i\rangle \|_2 \leq 2\epsilon, \tag{7.74}$$

where the second inequality follows from Equation (1.33), we have

$$\| B - I \|_\infty = \left\| \sum_{i=1}^{d} (|b_i'\rangle\langle b_i'| - |b_i\rangle\langle b_i|) \right\|_\infty \leq \sum_{i=1}^{d} \| |b_i'\rangle\langle b_i'| - |b_i\rangle\langle b_i| \|_\infty \leq 2d\epsilon, \tag{7.75}$$

and consequently $\left\| B^{-1/2} - I \right\|_\infty \leq \frac{2d\epsilon}{1-2d\epsilon}$. The latter can be seen by applying the definition of the spectral norm in terms of the singular values of its argument and showing that if $|x - 1| \leq y$ for $x \neq 0$, then $\left|\frac{1}{\sqrt{x}} - 1\right| \leq y/(1-y)$. We conclude that

$$\| |b\rangle - |b''\rangle \|_2 \leq \epsilon + \frac{2d\epsilon}{1-2d\epsilon} \leq \frac{(1+2d)\epsilon}{1-2d\epsilon} \leq (4d+2)\epsilon, \tag{7.76}$$

where the last inequality holds when $2d\epsilon \leq 1/2$.

With this bound in hand, we can now upper bound $\| U - U'' \|_\infty$ as

$$\begin{aligned}
\| U - U'' \|_\infty &= \left\| U^\dagger - (U'')^\dagger \right\|_\infty && (7.77)\\
&= \max_{|x\rangle \in \mathbb{C}^d \text{ s.t. } \| x \|_2 = 1} \left\| (U^\dagger - (U'')^\dagger)|x\rangle \right\|_2 && (7.78)\\
&\leq \sqrt{d} \max_{1 \leq i \leq d} |(\langle b_i| - \langle b_i''|)|x\rangle| && (7.79)\\
&\leq \sqrt{d}(4d+2)\epsilon && (7.80)\\
&\leq \delta, && (7.81)
\end{aligned}$$

where the first inequality follows since $\| x \|_2 \leq \sqrt{d} \| x \|_\infty$ for $|x\rangle \in \mathbb{C}^d$, the second inequality follows from the Cauchy-Schwarz inequality and Equation (7.76), and the third inequality from our definition of ϵ, as desired.

It remains to bound the cardinality of our δ-net: The number of different U'' in this net is at most $T_0 \leq (5/\epsilon)^{2d^2} = (30d^{3/2}/\delta)^{2d^2}$, since for each of the d columns of U'', we have at most $(5/\epsilon)^{2d}$ vectors in our pure state net N to choose from.

□

Lemma 7.14. *Let $S = \{|b_i\rangle\}_{i=1}^d \subseteq \mathbb{C}^d$ be an orthonormal basis for $\mathbb{C}^d$, and let $S' = \{|b'_i\rangle\}_{i=1}^d \subseteq \mathbb{C}^d$ satisfy $\| |b_i\rangle - |b'_i\rangle \|_2 \leq \epsilon$ for all i. Then if $\epsilon \in o(1/\sqrt{d})$, S' is a linearly independent set for large enough d.*

Proof. We proceed by contradiction. Assume S' is a linearly dependent set, i.e. there exist coefficients α_i, at least two of which are non-zero, such that

$$0 = \sum_i \alpha_i |b'_i\rangle = \sum_i \alpha_i (|b_i\rangle + |\epsilon_i\rangle). \tag{7.82}$$

for $|\epsilon_i\rangle \in \mathbb{C}^d$ with $\| |\epsilon_i\rangle \|_2 \leq \epsilon$. It follows that $\| \sum_i \alpha_i |b_i\rangle \|_2^2 = \| \sum_i \alpha_i |\epsilon_i\rangle \|_2^2$, implying

$$\sum_i |\alpha_i|^2 \leq \sum_{ij} |\alpha_i| \, |\alpha_j| \, |\langle \epsilon_i | \epsilon_j \rangle| \leq \epsilon^2 \sum_{ij} |\alpha_i| \, |\alpha_j| = \epsilon^2 \left(\sum_i |\alpha_i|^2 + \sum_{i \neq j} |\alpha_i| \, |\alpha_j| \right), \tag{7.83}$$

where the second inequality follows from the Cauchy-Schwarz inequality. Thus,

$$(1 - \epsilon^2) \sum_i |\alpha_i|^2 \leq \epsilon^2 \left(\sum_{i \neq j} |\alpha_i| \, |\alpha_j| \right) \leq \epsilon^2 \left(\sum_i |\alpha_i| \right)^2 \leq \epsilon^2 d \left(\sum_i |\alpha_i|^2 \right), \tag{7.84}$$

where the third inequality follows since $\| |v\rangle \|_1 \leq \sqrt{d} \| |v\rangle \|_2$ for any $|v\rangle \in \mathbb{C}^d$. Since S' is linearly dependent, $\sum_i |\alpha_i|^2 \neq 0$, and so dividing both end sides of the chain above by this quantity yields

$$1 \leq \epsilon^2 (d + 1), \tag{7.85}$$

which for $\epsilon \in o(1/\sqrt{d})$ yields a contradiction for large enough d. □

Acknowledgements for this chapter. We thank Fernando Brandão, Nicolas Brunner, Dagmar Bruß, Hermann Kampermann, Debbie Leung and Alexander Streltsov for helpful discussions.

Chapter 8

Characterizing quantumness via entanglement creation

This chapter is based on [110]:

S. Gharibian, M. Piani, G. Adesso, J. Calsamiglia and P. Horodecki. Characterizing quantumness via entanglement creation. *International Journal of Quantum Information*, 9(7 & 8):1701–1713, 2011, DOI: 10.1142/S0219749911008258,

In Chapter 8, we introduced an activation protocol which maps general non-classical (multipartite) correlations between given quantum systems into bipartite entanglement between the systems and an ancilla. Here, we study how this activation protocol can be used to entangle the starting systems themselves via entanglement swapping through a measurement on the ancilla. Furthermore, we bound the relative entropy of quantumness (a naturally arising measure of non-classicality in the scheme of Chapter 8) for a special class of separable states, the so-called classical-quantum states. In particular, we fully characterize the classical-quantum two-qubit states that are maximally non-classical.

8.1 Introduction and results

In this chapter, we continue our study of non-classical correlations (see Section 1.6.2 for a brief survey). Specifically, recall that the non-classicality of correlations present in multipartite quantum states is not due solely to the presence of entanglement. Namely, there

exist quantum states which are unentangled, but nevertheless exhibit traits that have *no* counterpart in the classical world. Such traits include no-local broadcasting [209] and the locking of correlations [87, 78] (see Section 1.6.2). Much effort has been devoted in recent years to characterize and quantify the non-classicality — or *quantumness* — of correlations [203, 138, 187, 119, 196, 118, 217, 185, 54, 209, 188, 94, 13, 216, 231] believed to be behind such feats.

In this context, we proposed an *activation* protocol in Chapter 7 which maps general non-classical (multipartite) correlations between input systems into bipartite entanglement between the systems and an ancilla. This was accomplished by letting the ancilla and input systems interact via CNOT gates with the systems acting as controls (see Section 7.3 for a formal description of the protocol). One advantange of this mapping is that it allows us to apply the tools and concepts of entanglement theory to the study of the quantumness of correlations. As an added bonus, the activation protocol, when considered in an adversarial context where the control bases are chosen so as to create the minimal amount of system-ancilla entanglement, provides an operational interpretation of the *relative entropy of quantumness* [54, 187, 118, 217, 196] as being the minimum distillable entanglement[210] *necessarily* (i.e. in the worst case scenario) created between the input systems and the ancilla.

In this chapter, we continue our study of the activation protocol of Chapter 7, and present two main contributions towards a better understanding of the quantumness of correlations.

Our Results: Here, we show the following.

1. Upper bounds on the non-classicality of separable states. We first give a non-trivial upper bound on the relative entropy of quantumness for a special class of separable states, the so-called *classical-quantum states* (see Section 1.6.2) (Lemma 8.1). Using this, we then fully characterize the classical-quantum two-qubit states which are maximally non-classical with respect to the relative entropy of quantumness (Lemma 8.2).

2. Entangling the input systems via entanglement swapping. The activation protocol of Chapter 7 demonstrates how to map non-classical correlations in an initial quantum system into entanglement between the system and an ancilla. However, one might prefer not to generate entanglement with an ancilla, but rather within the subsystems of the initial system itself.

We thus next study an approach for extending the activation protocol in order to entangle the input systems in such a manner as follows: We first run the original activation

protocol (i.e. we let each system interact with an ancilla). Next, we try to "swap"[152] the entanglement created between the input systems and ancilla back into entanglement among the input systems by performing a measurement on the ancilla alone. Note that we assume a worst-case scenario in performing this mapping: We ask, does there *exist* a choice of control bases for the activation protocol for which no entanglement can be created between the input systems with this approach, even if we allow post-selection after measuring the ancilla?

For this mapping, we derive conditions (Theorem 8.3, Corollary 8.4, discussion in Sections 8.4.2 and 8.4.3) under which entanglement can or cannot be swapped back into the input system. In particular, we find that there exist non-classical states which, despite *necessarily* leading to the creation of entanglement between systems and the ancilla in the activation protocol, may nevertheless fail to allow entanglement swapping back onto the initial system for a crafty choice of control bases.

Discussion and open questions. In this chapter, we first find bounds on the non-classicality (as measured by the relative entropy of quantumness) of classical-quantum states, and we characterize the maximally non-classical two-qubit classical-quantum states. It would be interesting to find bounds on the non-classicality of general separable states: from Chapter 7 we know that, for example, a separable state of two qubits can never be as non-classical as a maximally entangled pure state, but at present we do not know how large the gap between the two is. We remark that the maximally non-classical two-qubit CQ states found here (with respect to the relative entropy of quantumness) in Lemmas 8.1 and 8.2 match those found in Chapter 6 for the measure defined therein based on local unitary operations.

With respect to the swapping of the post-activation ancilla-system entanglement onto the original systems, we have both necessary conditions and sufficient conditions for the swapping to be possible in an adversarial scenario, but we lack conditions which are *simultaneously* necessary and sufficient. In finding such conditions, we suspect it would be beneficial to study the problem which arises in our swapping scheme: when is it possible to make a state entangled by rescaling rows and columns as in Equation (8.17)?

Finally, most of our results (e.g. Lemma 8.1, Theorem 8.3, and Corollary 8.4) apply to higher dimensional systems. However, it would be nice to extend Lemma 8.2 to this more general setting by characterizing the maximally non-classical classical-quantum states of higher dimension than qubits. Unfortunately, our approach here does not seem to apply in a straightforward manner to this setting, and further investigation is needed.

Organization of chapter. We begin in Section 8.2 with definitions and background information. In Section 8.3, we provide bounds on non-classicality for classical-quantum states, as measured by the relative entropy of quantumness. In Section 8.4, we present several results and observations regarding entangling input systems via the activation protocol and entanglement swapping.

8.2 Preliminaries

Throughout this chapter, we continue to use the notation and definitions from Chapter 7, which we briefly outline now. Recall from Definition 7.1 that a *strictly classically correlated* or *classical* quantum state ρ is one which diagonalizes in a local product basis $\mathcal{B}$.

We now outline the activation protocol of Chapter 7 (see Section 7.3 for further details). Consider an arbitrary state $\rho \in \mathcal{D}((\mathbb{C}^d)^{\otimes n})$ living in register A, where the ith local d-dimensional system lives in register A_i. We refer to A as the *system*. We further introduce a joint register A' of d-dimensional registers $A'_1, \ldots, A'_n$ of n ancilla qudit registers each initialized to the state $|0\rangle\langle 0|_{A'}$, henceforth called the *ancilla* (see Figure 7.1). The initial state of the joint $A : A'$ system is thus $\rho_{AA'} = \rho_A \otimes |0\rangle\langle 0|_{A'}^{\otimes n}$. For a given input ρ_A, we first consider for each i an adversarial application of a local unitary U_i to each A_i (i.e. this chooses the control basis for system i), and follow by applying one CNOT gate on each subsystem A_i (control qudit) and the corresponding ancillary party A'_i (target qudit). The final state of system plus ancilla at the end of this protocol is

$$\rho^f_{A:A'} = V(\rho_A \otimes |0\rangle\langle 0|_{A'}^{\otimes n})V^\dagger\,, \tag{8.1}$$

with $V = CNOT_{A:A'}(U_A \otimes I_{A'})$, $U_A = \otimes_{i=1}^n U_i$, and $CNOT_{AA'} = \bigotimes_{i=1}^n CNOT_{A_iA'_i}$. Recall that by Theorem 7.3, the output $\rho^f_{AA'}$ is separable across the $A : A'$ split if and only if ρ_A is classical. As done in Chapter 7, we henceforth refer to ρ_A and $\rho^f_{AA'}$ as ρ and ρ^f for simplicity, respectively.

It will be useful to also recall that ρ^f can be written as

$$\rho^f = \sum_{ij} \rho^{\mathcal{B}}_{ij} |i\rangle\langle j|_A \otimes |i\rangle\langle j|_{A'}, \tag{8.2}$$

where

$$\rho^{\mathcal{B}}_{ij} := \langle b(i)|\rho|b(j)\rangle, \tag{8.3}$$

for $|b(i)\rangle = U_A^\dagger|i\rangle$. In other words, ρ^f is of the maximally correlated [212] form in the $A : A'$ cut. Using this observation, we showed (Theorem 7.4) that if one quantifies the

minimum distillable entanglement generated across the $A : A'$ split in this protocol, the corresponding measure of non-classicality we obtain is given by

$$Q(\rho) = \min_{\mathcal{B}} \Big(S(\rho^{\mathcal{B}}) - S(\rho) \Big), \tag{8.4}$$

for $S(\rho)$ the von Neumann entropy, where the minimization is over all local product bases $\mathcal{B}$, and where $\rho^{\mathcal{B}} := \sum_i |b(i)\rangle\langle b(i)|\rho|b(i)\rangle\langle b(i)|$. It turned out (Corollary 7.5) that $Q(\rho)$ in fact coincides with the measure of non-classicality known as the *relative entropy of quantumness (REQ)* [54, 187, 118, 217, 196], bestowing the latter with an operational interpretation.

8.3 Upper bounds for separable states

In Theorem 7.6, we showed that for a bipartite state ρ_{AB} (where in the bipartite case we adopt the notational convention that $A_1 = A$ and $A_2 = B$), the quantity $Q(\rho_{AB})$ can achieve its maximum value of $\log d$ only for entangled states. We also showed that for increasing local dimension d, $Q(\rho_{AB})$ for certain separable ρ_{AB} can asymptotically approach $\log d$. What can be said, however, in the non-asymptotic setting? In other words, for a fixed local dimension d, how non-classical can separable ρ_{AB} be?

In this section, we first obtain a simple upper bound on $Q(\rho_{AB})$ for the subclass of separable states known as classical-quantum (CQ) states, that holds for arbitrary local dimensions. Recall from Section 1.6.2 that CQ states are those which can be written as $\rho_{AB} = \sum_{i=1}^{d_A} p_i |i\rangle\langle i| \otimes \rho_i$ for $\{|i\rangle\}_{i=1}^{d_A}$ an orthonormal basis, $\{p_i\}$ a probability distribution, and d_A and d_B the local dimensions of systems A and B. We then completely characterize the set of maximally non-classical two-qubit CQ states with respect to the relative entropy of quantumness $Q(\rho_{AB})$, and show that such states achieve $Q(\rho_{AB}) = 1/2$.

We begin with our claimed upper bound, which holds even when the local dimensions of A and B differ.

Lemma 8.1. *For any CQ state $\rho_{AB} = \sum_{i=1}^{d_A} p_i |i\rangle\langle i| \otimes \rho_i$, one has*

$$Q(\rho_{AB}) \leq \left(1 - \frac{1}{d_A}\right) \log_2 d_B. \tag{8.5}$$

Proof. We have

$$\begin{aligned} Q(\rho_{AB}) &= \min_{\mathcal{B}} S(\rho_{AB}^{\mathcal{B}}) - S(\rho_{AB}) && (8.6) \\ &= \min_{\mathcal{B}_B} S\left(\sum_{i=1}^{d_A} p_i |i\rangle\langle i| \otimes \left(\sum_{j=1}^{d_B} |b_B(j)\rangle\langle b_B(j)|\rho_i|b_B(j)\rangle\langle b_B(j)|\right)\right) - S(\rho_{AB}) \\ &= \min_{\mathcal{B}_B} \left(H(p) + \sum_{i=1}^{d_A} p_i S\left(\rho_i^{\mathcal{B}_B}\right)\right) - \left(H(p) + \sum_{i=1}^{d_A} p_i S(\rho_i)\right) && (8.7) \\ &= \min_{\mathcal{B}_B} \sum_{i=1}^{d_A} p_i \left[S\left(\rho_i^{\mathcal{B}_B}\right) - S(\rho_i)\right], && (8.8) \end{aligned}$$

where $H(p)$ denotes the Shannon entropy of the probability distribution $p = \{p_i\}_i$, and the second equality follows from choosing $\mathcal{B}_A$ to coincide with the basis $\{|i\rangle\}$. Let $p_m := \max_i p_i$. Our strategy is to let $\mathcal{B}_B$ project onto an eigenbasis of ρ_m, yielding:

$$\begin{aligned} Q(\rho_{AB}) &\leq \sum_{i \neq m} p_i \left[S\left(\rho_i^{\mathcal{B}_B}\right) - S(\rho_i)\right] && (8.9) \\ &\leq \sum_{i \neq m} p_i S\left(\rho_i^{\mathcal{B}_B}\right) && (8.10) \\ &\leq \left(1 - \frac{1}{d_A}\right) \log d_B, && (8.11) \end{aligned}$$

where the second inequality follows since $S(\rho_i) \geq 0$, and the third inequality follows since $p_m \geq 1/d_A$ and $S(\sigma_B) \leq \log d_B$ for any density operator σ_B. □

For a two-qubit CQ state ρ_{AB}, Lemma 8.1 implies $Q(\rho_{AB}) \leq 1/2$. We now show that this bound is tight by characterizing the set of CQ states attaining $Q(\rho_{AB}) = 1/2$.

Lemma 8.2. *Consider CQ state $\rho_{AB} \in \mathcal{D}(\mathbb{C}^2 \otimes \mathbb{C}^2)$ such that $\rho_{AB} = \sum_{i=1}^2 p_i |i\rangle\langle i| \otimes \rho_i$. Then $Q(\rho_{AB}) = 1/2$ if and only if $p_1 = p_2 = 1/2$ and $\rho_1 = |\psi_1\rangle\langle\psi_1|$ and $\rho_2 = |\psi_2\rangle\langle\psi_2|$ for some $|\psi_1\rangle, |\psi_2\rangle \in \mathbb{C}^2$ such that $|\langle\psi_1|\psi_2\rangle|^2 = 1/2$.*

Proof. That ρ_{AB} with $p_1 \neq 1/2$ implies $Q(\rho_{AB}) < 1/2$ follows immediately from Equation (8.10) and the fact that $0 \leq S(\sigma) \leq 1$ for any 1-qubit density operator σ. We thus henceforth assume $p_1 = p_2 = 1/2$. That ρ_1 and ρ_2 must be pure now also follows analogously, for if, say, ρ_1 is mixed, then we simply choose $\mathcal{B}_B$ in Equation (8.9) to instead project onto an eigenbasis of ρ_2, and use the fact that $S(\rho_1) > 0$ to achieve $Q(\rho_{AB}) < 1/2$.

We thus henceforth assume $\rho_1 = |\psi_1\rangle\langle\psi_1|$ and $\rho_2 = |\psi_2\rangle\langle\psi_2|$ for some $|\psi_1\rangle, |\psi_2\rangle \in \mathbb{C}^2$. It remains to show that we must have $|\langle\psi_1|\psi_2\rangle|^2 = 1/2$.

Plugging ρ_{AB} into Equation (8.8) and noting that $S(\rho_1) = S(\rho_2) = 0$, we have

$$\begin{aligned} Q(\rho_{AB}) &= \frac{1}{2}\min_{\mathcal{B}_B}\left[S([|\psi_1\rangle\langle\psi_1|]^{\mathcal{B}_B}) + S([|\psi_2\rangle\langle\psi_2|]^{\mathcal{B}_B})\right] \qquad (8.12)\\ &= \frac{1}{2}\min_{\mathcal{B}_B}\left[H\left(|\langle b_B(0)|\psi_1\rangle|^2, |\langle b_B(1)|\psi_1\rangle|^2\right) + H\left(|\langle b_B(0)|\psi_2\rangle|^2, |\langle b_B(1)|\psi_2\rangle|^2\right)\right]\\ &= \frac{1}{2}\min_{|b_B(0)\rangle}\left[H\left(|\langle b_B(0)|\psi_1\rangle|^2, \left|\langle b_B(0)|\psi_1^\perp\rangle\right|^2\right) + \right.\\ &\qquad \left. H\left(|\langle b_B(0)|\psi_2\rangle|^2, \left|\langle b_B(0)|\psi_2^\perp\rangle\right|^2\right)\right], \qquad (8.13)\end{aligned}$$

where $\langle\psi_1|\psi_1^\perp\rangle = \langle\psi_2|\psi_2^\perp\rangle = 0$, and where the last equality follows since $|b_B(j)\rangle\langle b_B(j)|$ are rank-one projectors. Note that one can think of the last equality as effectively switching the roles of the measurement and the target state, so that the minimization can be thought of as being taken over all pure *target* states $|b_B(0)\rangle$ with respect to *measurements* in the bases $\mathcal{B}_1 := \left\{|\psi_1\rangle, \left|\psi_1^\perp\right\rangle\right\}$ and $\mathcal{B}_2 := \left\{|\psi_2\rangle, \left|\psi_2^\perp\right\rangle\right\}$. We can now plug Equation (8.13) into the well-known entropic uncertainty relation of Maassen and Uffink [189, 251], which states that for classical distributions P_c and P_d obtained by measuring pure state $|\psi\rangle$ with respect to orthonormal bases $\mathcal{C} = \{|c\rangle\}$ and $\mathcal{D} = \{|d\rangle\}$, respectively, we have

$$\frac{1}{2}(H(P_c) + H(P_d)) \geq -\log f(\mathcal{C}, \mathcal{D}), \qquad (8.14)$$

where $f(\mathcal{C}, \mathcal{D}) := \max\{|\langle c|d\rangle| \mid |c\rangle \in \mathcal{C}, |d\rangle \in \mathcal{D}\}$. We thus obtain:

$$Q(\rho_{AB}) \geq \max_{|\phi_1\rangle\in\mathcal{B}_1, |\phi_2\rangle\in\mathcal{B}_2} -\log|\langle\phi_1|\phi_2\rangle|. \qquad (8.15)$$

Note that this lower bound attains its maximum value of $1/2$ if $\mathcal{B}_1$ and $\mathcal{B}_2$ are mutually unbiased, i.e. when $|\langle\psi_1|\psi_2\rangle|^2 = 1/2$. On the other hand, suppose $\mathcal{B}_1$ and $\mathcal{B}_2$ are *not* mutually unbiased, i.e. suppose without loss of generality that $|\langle\psi_1|\psi_2\rangle|^2 > 1/2$. Then choosing $|b_B(0)\rangle = |\psi_1\rangle$ in Equation (8.13) yields $Q(\rho_{AB}) < 1/2$. The claim follows. □

Combining Lemmas 8.1 and 8.2, we obtain a characterization of the set of two-qubit CQ states which are deemed maximally non-classical by Q. Such states include, for example, the CQ state

$$\rho = \frac{1}{2}|0\rangle\langle 0| \otimes |0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| \otimes |+\rangle\langle +| = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & \frac{1}{2} & \frac{1}{2}\\ 0 & 0 & \frac{1}{2} & \frac{1}{2}\end{pmatrix}, \qquad (8.16)$$

where $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$.

8.4 Swapping the ancilla-system entanglement onto the system

We now explore the possibility of generating entanglement in the *original* system AB by projecting the ancilla systems A' of the state ρ^f of (8.1) jointly onto an entangled pure state. In other words, we consider an entanglement swapping process [152] that maps the system-ancilla entanglement onto the systems AB. As we are only interested in knowing whether this is possible (rather than, say, in the probability of success), the filtering via a pure state is not restrictive and corresponds to the best possible strategy. Our results indicate that this feat is possible for some, but not all, separable non-classical states.

We begin by noting that thanks to the maximally-correlated form of ρ^f (Equation (8.2)), we have that the (unnormalized) final state of system AB after projecting the ancilla system onto (normalized) state $|\phi\rangle = \sum_i \alpha_i |i\rangle \in (\mathbb{C}^d)^{\otimes n}$ is given by

$$\rho_\phi := \mathrm{Tr}_{A'}(\rho^f |\phi\rangle\langle\phi|_{A'}) = \sum_{ij} \left[\rho^{\mathcal{B}}_{ij} \alpha_i \alpha_j^*\right] |i\rangle\langle j|, \tag{8.17}$$

with $\rho^{\mathcal{B}}_{ij}$ defined in Equation (8.3). Hence, the resulting (unnormalized) state ρ_ϕ is simply the Hadamard product of the original state (represented in the $\mathcal{B}$ basis) and $|\phi\rangle\langle\phi|$ (represented in the computational basis), i.e. $\rho_\phi = \rho^{\mathcal{B}} \circ |\phi\rangle\langle\phi|$ for $\circ$ the Hadamard product defined such that $(C \circ D)(i,j) := C(i,j)D(i,j)$.

As previously mentioned, our goal is to answer the following question: For a given input ρ, is it true that for *any* choice of starting local bases for the CNOT gates in the activation protocol, there exists a state $|\phi\rangle$ such that ρ_ϕ is entangled (across its constituent local d-dimensional systems)?

In Section 8.4.1 we provide a simple sufficient condition under which the generation of entanglement in the original system is always possible with an appropriate choice of $|\phi\rangle$, regardless of the choice of adversarial local unitary. We then observe that this condition holds for all pseudo-isotropic states as in Equation (8.19), with $|\psi\rangle$ entangled and $p > 0$. In Sections 8.4.2 and 8.4.3, we provide examples of classical-quantum (CQ) and quantum-quantum (QQ) separable states, respectively, for which entanglement in AB cannot be generated in this fashion, i.e. there exists a choice of U_{AB} that prevents the generation of entanglement in AB via the swapping of system-ancilla entanglement, *even if* there is

necessarily entanglement between the system-ancilla cut after the activation protocol is run.

8.4.1 Sufficient condition for entanglement swapping

We focus again on the bipartite case $A_1 = A$, $A_2 = B$. We have the following simple condition which ensures the swapping of entanglement is possible.

Theorem 8.3. *If for any choice of local basis $\mathcal{B}$, there exists a non-zero off-diagonal element of an off-diagonal block of $\rho_{AB}^{\mathcal{B}}$, i.e. if for all $\mathcal{B} = \mathcal{B}_A\mathcal{B}_B$ there exists a choice of $i \neq j$ and $k \neq l$ such that $\langle b_A(i)b_B(k)|\rho_{AB}|b_A(j)b_B(l)\rangle \neq 0$, then it is possible to swap entanglement back into the input systems (regardless of the choice of $\mathcal{B}$), i.e. there exists a $|\phi\rangle$ such that ρ_ϕ is entangled.*

Proof. The strategy of the proof is to choose $|\phi\rangle$ so that the result of the Hadamard product in Equation (8.17) is non-positive under partial transposition (NPT) [206, 146]. Fix any choice of local basis $\mathcal{B}$. By assumption, we know there exist indices $i \neq j$ and $k \neq l$ such that $\langle b_A(i)b_B(k)|\rho_{AB}|b_A(j)b_B(l)\rangle \neq 0$. In order to ensure that ρ_ϕ is NPT, we thus choose $|\phi\rangle$ to single out these non-zero off-diagonal terms by setting

$$|\phi\rangle = \frac{1}{\sqrt{2}}(|ik\rangle + |jl\rangle). \tag{8.18}$$

With this choice of $|\phi\rangle$, ρ_ϕ becomes a Hermitian matrix with only four non-zero entries, two of which lie on the diagonal at positions $|i\rangle\langle i| \otimes |k\rangle\langle k|$ and $|j\rangle\langle j| \otimes |l\rangle\langle l|$, and two of which lie at off-diagonal positions of off-diagonal blocks at $|i\rangle\langle j| \otimes |k\rangle\langle l|$ and $|j\rangle\langle i| \otimes |l\rangle\langle k|$ (i.e. the four entries form the four corners of a square). It follows that the partial transpose of ρ_ϕ is not positive. □

Corollary 8.4. *For any*

$$\rho(\psi, p)_{AB} := (1-p)\frac{I_{AB}}{D} + p|\psi\rangle\langle\psi|_{AB}, \tag{8.19}$$

with I_{AB}/D the maximally mixed state for AB and D the dimension of AB, if $|\psi\rangle$ is entangled and $p > 0$, then there exists a choice of $|\phi\rangle$ such that ρ_ϕ is entangled.

Proof. Since the maximally mixed component of (8.19) is diagonal with respect to any choice of local bases, it suffices to argue that $|\psi\rangle$ satisfies the condition of Theorem 1. This easily follows from the fact $|\psi\rangle$ is entangled, and thus has, up to local unitaries, a Schmidt decomposition $\sum_{k=0}^{d_A-1}\sqrt{\lambda_k}|k\rangle|k\rangle$, with $\lambda_0 \geq \lambda_1 > 0$. □

Corollary 8.4 shows that for any value of $p > 0$, entanglement can be transferred to the original system for the pseudo-isotropic state $\rho(p, \psi)$ of Equation (8.19), even for values of p which correspond to *separable* states (recall that for p small enough, the state $\rho(p, \psi)$ is separable due to the existence of a separable ball around the maximally mixed state [153, 124]). We remark that for all $p > 0$ and entangled $|\psi\rangle$, $\rho(p, \psi)$ is known to be non-classical [118], and so here the non-classicality of the starting state allows us to create entanglement in the original systems AB by applying the activation protocol followed by our entanglement swapping procedure.

8.4.2 Classical-quantum separable states

In Section 8.4.1, we demonstrated that for certain non-classically correlated states, entanglement can be mapped back into the original system after the activation protocol is run. Can this be achieved with *any* type of non-classically correlated input ρ? We now show that the answer is no — there exist separable non-classical ρ such that, while entanglement is always generated in the activation protocol between systems and ancilla independently of the local unitaries U_A and U_B, a proper adversarial choice of local unitaries U_A and U_B can nevertheless prevent entanglement from being mapped back to the system.

Consider the separable non-classical CQ state of Equation (8.16). By Equation (8.17), note that when the adversarial local unitaries are chosen as $U_A = U_B = I$, we have

$$\rho_\phi = \rho \circ |\phi\rangle\langle\phi|. \tag{8.20}$$

Since ρ is block diagonal, it hence follows that ρ_ϕ is block diagonal, since the Hadamard product cannot change this block diagonal structure regardless of the choice of $|\phi\rangle$. We conclude that there exists a choice of local bases (i.e the computational basis) with respect to which ρ_ϕ is always separable for all $|\phi\rangle$, i.e. it is not possible to project the (necessarily present) system-ancilae entanglement generated in the activation protocol back onto the system. In fact, this proof approach holds for *any* CQ (or QC) state that is not strictly classically correlated, implying that for such states, there is a choice of local unitaries for which, even if entanglement is created between system and ancilla in the activation protocol, such entanglement cannot be swapped back into the input system.

8.4.3 Quantum-quantum separable states

Based on the results in Section 8.4.2, one might hope that entanglement generation in separable starting systems is possible if ρ is not CQ nor QC (i.e. ρ is what we might call

QQ separable). We provide a counterexample to this conjecture here — namely, we show that there exist QQ separable states for which an adversarial choice of local bases in the activation protocol prevents the swapping of ancilla-system entanglement back into the input systems.

To do so, consider the separable QQ operator:

$$\rho_{AB} = \frac{1}{2}|0\rangle\langle 0| \otimes |+\rangle\langle +| + \frac{1}{2}|+\rangle\langle +| \otimes |0\rangle\langle 0| = \frac{1}{4}\begin{pmatrix} 2 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{8.21}$$

To prove our claim, as in Section 8.4.2, we choose local adversarial unitaries $U_A = U_B = I$ and show that $\rho^f = \rho_{AB} \circ |\phi\rangle\langle\phi|$ is separable for any choice of $|\phi\rangle$. The latter is shown by first deriving a condition under which the eigenvalues of Hermitian operators with a structure similar to ρ remain invariant under partial transposition. We then show that ρ fulfills this condition for any choice of $|\phi\rangle$, implying ρ always remains separable, since the partial transpose is a necessary and sufficient condition for separability of two-qubit states [146].

Lemma 8.5. *Given any Hermitian operator X acting on $\mathbb{C}^2 \otimes \mathbb{C}^2$ with off-diagonal blocks which are diagonal, i.e.*

$$X = \begin{pmatrix} a_{11} & a_{12} & a_{13} & 0 \\ a_{12}^* & a_{22} & 0 & a_{24} \\ a_{13}^* & 0 & a_{33} & a_{34} \\ 0 & a_{24}^* & a_{34}^* & a_{44} \end{pmatrix}, \tag{8.22}$$

if either $a_{12}a_{34}^* \in \mathbb{R}$ or $a_{13}a_{24}^* \in \mathbb{R}$, then the spectrum of A is invariant under partial transposition.

Proof. Let $p_X(\lambda)$ and $p_{X^\Gamma}(\lambda)$ denote the characteristic polynomials of X and X^Γ, the partial transpose of X, respectively. Then

$$p_X(\lambda) - p_{X^\Gamma}(\lambda) = 2\,\mathrm{Re}(a_{12}a_{34}^* a_{13}a_{24}^* - a_{12}^* a_{34} a_{13} a_{24}^*) = 4\,\mathrm{Im}(a_{13}^* a_{24})\,\mathrm{Im}(a_{12}a_{34}^*), \tag{8.23}$$

where $\mathrm{Re}(x)$ ($\mathrm{Im}(x)$) denotes the real (imaginary) part of x. The claim follows for $a_{12}a_{34}^* \in \mathbb{R}$. An analogous calculation yields the $a_{13}a_{24}^* \in \mathbb{R}$ case. □

With Lemma 8.5 in hand, it is easy to see that ρ^f has a positive partial transpose (and is hence separable) for all $|\phi\rangle$ — specifically, we observe that ρ satisfies the conditions of

Lemma 8.5 since $a_{12}a_{34}^* = (1/4)(0) = 0$, and this in particular holds even after taking the Hadamard product with any $|\phi\rangle\langle\phi|$. Since ρ is positive semidefinite, it thus follows from Lemma 8.5 that ρ^f must also be positive semidefinite under partial transposition and hence separable. Thus, there exist QQ separable states for which system-ancilla entanglement cannot be mapped back to the system.

Theorem 8.3 tells us that if a two-qubit state ρ has off-diagonal terms on its off-diagonal blocks for any choice of local bases, then entanglement can be created among the systems via swapping. On the other hand, if ρ is restricted to having off-diagonal blocks which are diagonal, as was seen with the CQ and QQ counterexamples considered in Sections 8.4.2 and 8.4.3, then there are choices of local initial rotations such that entanglement generation among the systems is not necessarily possible (actually, in the CQ case, entanglement generation is not possible for *any* choice of local initial rotations).

One could ask whether this "diagonal off-diagonal" block structure is sufficient to rule out the possibility of entanglement generation. The answer is negative. Consider the following (un-normalized) positive semidefinite operator which has diagonal off-diagonal blocks:

$$\rho = \begin{pmatrix} \frac{3}{2} & i & 1 & 0 \\ -i & \frac{3}{2} & 0 & i \\ 1 & 0 & \frac{3}{2} & 1 \\ 0 & -i & 1 & \frac{3}{2} \end{pmatrix}. \tag{8.24}$$

It turns out that the partial transposition of ρ has a negative eigenvalue (observe that ρ thus also necessarily violates the conditions of Lemma 8.5). Hence, despite the fact that ρ has off-diagonal blocks which are diagonal, it is nevertheless entangled, implying entanglement transfer to the system is possible for any choice of local bases: indeed, the Hadamard product can be chosen to be trivial, so that the projection simply gives back (a locally rotated and unnormalized) ρ_A.

Chapter 9

Conclusion

In this thesis, we have studied three areas in quantum computation and information: the approximability of quantum problems, quantum proof systems, and non-classical correlations. Our results in each of these areas are summarized as follows.

With respect to approximation, we have completed some of the first works in an area aiming to understand the computational complexity of efficiently and rigorously computing approximate solutions to problems which are complete for quantum complexity classes. In Chapter 2, we demonstrated a polynomial time approximation algorithm for dense instances of the canonical QMA-complete problem, the local Hamiltonian problem. This required the derivation of a lower bound on the approximation ratio achievable by product state assignments, which as discussed in Chapter 2, can be seen as negative progress towards a sought-after quantum PCP theorem. Among other open questions discussed therein, perhaps the most natural direction here is the pursuit of further new approximation algorithms for problems complete for QMA, the quantum generalization of NP. In Chapter 3, we then proceeded in the opposite direction by demonstrating hardness of approximation results for a new quantum complexity class we defined, cq-Σ_2. This class is an arguably natural generalization of Σ_2^p, and the hard-to-approximate problems for cq-Σ_2 we considered are generalizations of classical covering problems obtained via the notion of *quantum* constraint satisfaction (i.e. local Hamiltonian constraints). Aside from the obvious open questions here regarding further hardness of approximation results such as a quantum PCP theorem, we would be interested to know to what extent the class cq-Σ_2 itself may play an important role in quantum complexity theory, just as Σ_2^p has proven valuable in the classical setting.

With respect to quantum proof systems, in Chapter 4 we focused on the question of

whether multiple unentangled provers can be simulated by a single prover in the context of QMA proof systems. As the question of whether two provers are as good as one (i.e. is QMA = QMA(2)?) remains open despite much effort, we focused our attention on variants of QMA in which the verification protocol is suitably restricted. In this setting, we showed various results, including a collapse to QMA for a restricted variant of QMA(poly), and an alternate proof of a parallel repetition theorem for SepQMA(2). Understanding the non-trivial "power of unentanglement" [10] between quantum provers remains a challenging and interesting direction of work.

Finally, with respect to non-classical correlations, in Chapter 5, we first motivated the study of such correlations by examining their role in the DQC1 trace estimation algorithm, as well as the quantum communication task of locking. Above all, understanding the precise role such correlations play in mixed-state quantum computing remains an important open question. In Chapter 6, we then proposed a novel scheme for quantifying non-classical correlations based on a special class of local unitary operations. This raised the question as to how the notions of "disturbance under measurement" and "disturbance under unitary operations" differ in their characterization of non-classical correlations. Finally, Chapters 7 and 8 introduced and studied a protocol which "activates" non-classical correlations present in a multipartite quantum system into entanglement between the system and an ancilla. Aside from yielding a new framework through which new non-classicality measures can be discovered, our study here also revealed a surprising result: That mixedness in quantum states can play a very important role in giving rise to non-classical correlations, both for separable and entangled states. We would be interested to see how this framework may be further developed, and moreover whether the ideas behind it may prove useful in a quantum computational setting.

In conclusion, the field of quantum computation and information is, after over two decades of study, arguably no longer in its infancy. With a solid theoretical base and formalism in place, including the quantum circuit model, quantum complexity classes and proof systems, and foundations for quantum information theory, the field now covers a large number of areas of study, of which our focus here is but a small part. Yet, whether quantum computers will, at a practical level, indeed be the wave of the future, is in our opinion not yet entirely clear. What *is* clear, however, is that no matter the outcome, the lessons learned through this line of work have already taught us much about the physical world around us. Indeed, the study of this field has united the physics and computer science communities towards a common ultimate goal: To probe the physical limits of nature and computing themselves. This in itself is no small feat. As it stands, information *is* physical. We would not (and *could not*) have it any other way.

References

[1] List of Intel microprocessors. http://en.wikipedia.org/wiki/List_of_Intel_microprocessors#Original_Pentium.

[2] One small step for Cal, a quantum leap for mankind. http://sciencereview.berkeley.edu/read/fall-2011/one-small-step-for-cal-a-quantum-leap-for-mankind/.

[3] Quantum processor wierdness. http://www.ctoedge.com/content/quantum-processor-weirdness.

[4] Ramones - Rock n' Roll Hall of Fame induction (March 2002). http://www.youtube.com/watch?v=BZEEaXJar10, beginning at 2:47.

[5] Talk: Albert Einstein. http://en.wikiquote.org/wiki/Talk:Albert_Einstein.

[6] S. Aaronson. The quantum PCP manifesto, 2006. http://scottaaronson.com/blog/?p=139.

[7] S. Aaronson. On perfect completeness for QMA. *Quantum Information & Computation*, 9(1 & 2), 2009.

[8] S. Aaronson. BQP and the polynomial hierarchy. In *Proceedings of the 42nd ACM Symposium on the Theory of Computing (STOC 2010)*, pages 141–150, 2010.

[9] S. Aaronson. A counterexample to the generalized Linial-Nisan conjecture. Available at arXiv.org e-Print quant-ph/1110.6126v1, 2011.

[10] S. Aaronson, S. Beigi, A. Drucker, B. Fefferman, and P. Shor. The power of unentanglement. *Theory of Computing*, 5:1–42, 2009.

[11] S. Aaronson and G. Kuperberg. Quantum versus classical proofs and advice. *Theory of Computing*, 3:129–157, 2007.

[12] S. Aaronson, G. Kuperberg, and C. Granade. Complexity Zoo. http://qwiki.stanford.edu/index.php/Complexity_Zoo.

[13] G. Adesso and A. Datta. Quantum versus classical correlations in Gaussian states. *Physical Review Letters*, 105:030501, 2010.

[14] L. Adleman, J. DeMarrais, and M. Huang. Quantum computability. *SIAM Journal on Computing*, 26(5):1524–1540, 1997.

[15] D. Aharanov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev. Adiabatic quantum computation is equivalent to standard quantum computation. *SIAM Journal on Computing*, 37(1):166–194, 2007.

[16] D. Aharonov, I. Arad, and S. Irani. Efficient algorithm for approximating one-dimensional ground states. *Physical Review A*, 82:012315, 2010.

[17] D. Aharonov, I. Arad, Z. Landau, and U. Vazirani. The detectibility lemma and quantum gap amplification. In *Proceedings of 41st ACM Symposium on Theory of Computing (STOC 2009)*, volume 287, pages 417–426, 2009.

[18] D. Aharonov, M. Ben-Or, F. Brandão, and O. Sattath. The pursuit for uniqueness: Extending Valiant-Vazirani theorem to the probabilistic and quantum settings. Available at arXiv.org e-Print quant-ph/0810.4840v1, 2008.

[19] D. Aharonov and L. Eldar. On the complexity of commuting local Hamiltonians, and tight conditions for Topological Order in such systems. In *Proceedings of the 52nd IEEE Symposium on Foundations of Computer Science (FOCS 2011)*, pages 334–343, 2011.

[20] D. Aharonov, D. Gottesman, S. Irani, and J. Kempe. The power of quantum systems on a line. *Communications in Mathematical Physics*, 287:41–65, 2009.

[21] D. Aharonov, A. Kitaev, and N. Nisan. Quantum circuits with mixed states. In *Proceedings of 13th ACM Symposium on Theory of Computing (STOC 1998)*, pages 20–30, 1998.

[22] D. Aharonov and T. Naveh. Quantum NP - A survey. Available at arXiv.org e-Print quant-ph/0210077v1, 2002.

[23] N. Alon, W. F. de la Vega, R. Kannan, and M. Karpinski. Random sampling and approximation of MAX-CSP problems. In *Proceedings of the 34th Symposium on Theory of Computing (STOC 2002)*, pages 232–239, 2002.

[24] A. Ambainis, A. M. Childs, B. W. Reichardt, R. Spalek, and S. Zhang. Any AND-OR formula of size N can be evaluated in time $N^{1/2+o(1)}$ on a quantum computer. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science (FOCS 2007)*, pages 363–372, 2007.

[25] A. Ambainis, L. J. Schulman, and U. V. Vazirani. Computing with highly mixed states. In *Proceedings of the 32nd ACM Symposium on Theory of Computing*, pages 697–704, 2000.

[26] I. Arad. A note about a partial no-go theorem for quantum PCP. Available at arXiv.org e-Print quant-ph/1012.3319, 2010.

[27] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.

[28] S. Arora, D. Karger, and M. Karpinski. Polynomial time approximation schemes for dense instances of NP-hard problems. *Journal of Computer and System Sciences*, 58:193–210, 1999.

[29] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998. Prelim. version FOCS '92.

[30] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998. Prelim. version FOCS '92.

[31] K. M. R. Audenaert. A sharp Fannes-type inequality for the von Neumann entropy. *Journal of Physics A*, 40:8127–8136, 2006.

[32] P. Austrin and E. Mossel. Approximation resistant predicates from pairwise independence. In *Proceedings of the 23rd IEEE Conference on Computational Complexity*, pages 249–258, 2008.

[33] L. Babai. Trading group theory for randomness. In *Proceedings of 17th ACM Symposium on Theory of Computing (STOC 1985)*, pages 421–429, 1985.

[34] M. Ballester and S. Wehner. Entropic uncertainty relations and locking: tight bounds for mutually unbiased bases. *Physical Review A*, 75:022319, 2007.

[35] N. Bansal, S. Bravyi, and B. M. Terhal. Classical approximation schemes for the ground-state energy of quantum and classical Ising spin Hamiltonians on planar graphs. *Quantum Information & Computation*, 9(7&8):0701–0720, 2009.

[36] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. A generalized no-broadcasting theorem. *Physical Review Letters*, 99:240501, 2007.

[37] H. Barnum, C. M. Caves, C. A. Fuchs, R. Josza, and B. Schumacher. Noncommuting mixed states cannot be broadcast. *Physical Review Letters*, 76:2818–2821, 1996.

[38] C. Bazgan, W. F. de la Vega, and M. Karpinski. Polynomial time approximation schemes for dense instances of minimum constraint satisfaction. *Random Structures & Algorithms*, 23(1):73–91, 2003.

[39] S. Beigi. NP vs $\mathrm{QMA}_{\log}(2)$. *Quantum Information & Computation*, 10:0141–0151, 2010.

[40] S. Beigi and P. W. Shor. On the complexity of computing zero-error and Holevo capacity of quantum channels. Available at arXiv.org e-Print quant-ph/0709.2090v3, 2007.

[41] M. Bellare, O. Goldreich, and S. Goldwasser. Randomness in interactive proofs. *Computational Complexity*, 3:319–354, 1993.

[42] M. Bellare and J. Rompel. Randomness-efficient oblivious sampling. In *Proceedings of the 28th Annual IEEE Symposium on the Foundations of Computer Science (FOCS 1994)*, pages 276–287, 1994.

[43] P. Benioff. The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines. *Journal of Statistical Physics*, 22:563–591, 1980.

[44] P. Benioff. Quantum mechanical Hamiltonian models of Turing machines. *Journal of Statistical Physics*, 29:515–546, 1982.

[45] P. Benioff. Quantum mechanical Hamiltonian models of Turing machines that dissipate no energy. *Physical Review Letters*, 48:1581–1585, 1982.

[46] E. Bernstein and U. Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997.

[47] E. Biham, G. Brassard, D. Kenigsberg, and T. Mor. Quantum computing without entanglement. *Theoretical Computer Science*, 320:15, 2004.

[48] H. Blier and A. Tapp. All languages in NP have very short quantum proofs. In *Proceedings of the 3rd International Conference on Quantum, Nano and Micro Technologies*, pages 34–37, 2009. Available at arXiv.org e-Print quant-ph/0709.0738v2, first posted in 2007.

[49] S. Boixo, L. Aolita, D. Cavalcanti, K. Modi, and A. Winter. Quantum locking of classical correlations and quantum discord of classical-quantum states. *International Journal of Quantum Information*, 9:1643–1651, 2011.

[50] M. Born and A. Einstein. *The Born-Einstein letters: correspondence between Albert Einstein and Max and Hedwig Born from 1916–1955*. Walker, 1971.

[51] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press.

[52] F. Brandão. *Entanglement Theory and the Quantum Simulation of Many-Body Physics*. PhD thesis, Imperial College London, London, 2008. Available at arXiv.org e-Print quant-ph/1011.2751v2.

[53] F. Brandão, M. Christandl, and J. Yard. A quasipolynomial-time algorithm for the quantum separability problem. In *Proceedings of the 43rd ACM Symposium on Theory of Computing (STOC 2011)*, pages 343–351, 2011.

[54] S Bravyi. Entanglement entropy of multipartite pure states. *Physical Review A*, 67(1):012313, 2003.

[55] S. Bravyi. Efficient algorithm for a quantum analogue of 2-SAT. Available at arXiv.org e-Print quant-ph/0602108v1, 2006.

[56] S. Bravyi, A. Bessen, and B. Terhal. Merlin-Arthur games and stoquastic complexity. Available at arXiv.org e-Print quant-ph/0611021v2, 2006.

[57] S. Bravyi, D. DiVincenzo, and D. Loss. Polynomial-time algorithm for simulation of weakly interacting quantum spin systems. *Communications in Mathematical Physics*, 287:41–65, 2009.

[58] S. Bravyi, D. DiVincenzo, R. Oliveira, and B. Terhal. The complexity of stoquastic local Hamiltonian problems. *Quantum Information & Computation*, 8(5):0361–0385, 2008.

[59] S. Bravyi and B. Terhal. Complexity of stoquastic frustration-free Hamiltonians. *SIAM Journal on Computing*, 39(4):1462, 2009.

[60] S. Bravyi and M. Vyalyi. Commutative version of the local Hamiltonian problem and common eigenspace problem. *Quantum Information & Computation*, 5(3):187–215, 2005.

[61] D. Bruß. Characterizing entanglement. *Journal of Mathematical Physics*, 43:4237, 2001.

[62] D. Cavalcanti, L. Aolita, S. Boixo, K. Modi, M. Piani, and A. Winter. Operational interpretations of quantum discord. *Physical Review A*, 83:032324, 2011.

[63] M. Charikar, K. Makarychev, and Y. Makarychev. Near-optimal algorithms for maximum constraint satisfaction problems. In *Lecture Notes in Computer Science*, volume 4627, pages 149–163, 2007.

[64] J. Chen and A. Drucker. Short multi-prover quantum proofs for SAT without entangled measurements. Available at arXiv.org e-Print quant-ph/1011.0716v2, 2010.

[65] A. Chiesa and M. Forbes. Improved soundness for QMA with multiple provers. Available at arXiv.org e-Print quant-ph/1108.2098v1, 2011.

[66] A. M. Childs, R. Cleve, S. P. Jordan, and D. Yeung. Discrete-query quantum algorithm for NAND trees. *Theory of Computing*, 5:119–123, 2009.

[67] E. Chitambar. Quantum correlations in large-dimensional states of high symmetry. Available at arXiv.org e-Print quant-ph/1110.3057, 2011.

[68] T. K. Chuan, J. Maillard, K. Modi, T. Paterek, M. Paternostro, and M. Piani. Role of quantumness of correlations in entanglement distribution. Available at arXiv.org e-Print quant-ph/1203.1268v2, 2012.

[69] J. I. Cirac and F. Verstraete. Renormalization and tensor product states in spin chains and lattices. *Journal of Physics A*, 42(50):504004, 2009.

[70] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23:880, 1969.

[71] R. Cleve, W. Slofstra, F. Unger, and S. Upadhyay. Perfect parallel repetition theorem for quantum XOR proof systems. *Computational Complexity*, 17(2):282–299, 2008.

[72] S. Cook. The complexity of theorem proving procedures. In *Proceedings of the 3rd ACM Symposium on Theory of Computing (STOC 1972)*, pages 151–158, 1972.

[73] T. S. Cubitt, F. Verstraete, W. Dür, and J. I. Cirac. Separable states can be used to distribute entanglement. *Physical Review Letters*, 91:037902, 2003.

[74] B. Dakić, Y. Ole Lipp, X. Ma, M. Ringbauer, S. Kropatschek, S. Barz, T. Paterek, V. Vedral, A. Zeilinger, Č. Brukner, and P. Walther. Quantum discord as optimal resource for quantum communication. Available at arXiv.org e-Print quant-ph/1203.1629v1, 2012.

[75] B. Dakić, V. Vedral, and Č. Brukner. Necessary and sufficient condition for nonzero quantum discord. *Physical Review Letters*, 105(190502), 2010.

[76] A. Datta. A condition for nullity of quantum discord. Available at arXiv.org e-Print quant-ph/1003.5256, 2010.

[77] A. Datta, S. T. Flammia, and C. M. Caves. Entanglement and the power of one qubit. *Physical Review A*, 72:042316, 2005.

[78] A. Datta and S. Gharibian. Signatures of nonclassicality in mixed-state quantum computation. *Physical Review A*, 79:042325, 2009. DOI: 10.1103/PhysRevA.79.042325, © 2009 American Physical Society, pra.aps.org.

[79] Animesh Datta. *Studies on the Role of Entanglement in Mixed-state Quantum Computation*. PhD Thesis, University of New Mexico, 2008. Available at arXiv:0807.4490v1.

[80] Animesh Datta, Anil Shaji, and Carlton M. Caves. Quantum discord and the power of one qubit. *Physical Review Letters*, 100:050502, 2008.

[81] W. F. de la Vega. MAX-CUT has a randomized approximation scheme in dense graphs. *Random Structures & Algorithms*, 8(3):187–198, 1996.

[82] W. F. de la Vega and M. Karpinski. Polynomial time approximation of dense weighted instances of MAX-CUT. *Random Structures & Algorithms*, 16:314–332, 2000.

[83] W. F. de la Vega, M. Karpinski, R. Kannan, and S. Vempala. Tensor decomposition and approximation schemes for constraint satisfaction problems. In *Proceedings of the 37th Symposium on Theory of Computing (STOC 2005)*, pages 747–754. ACM Press, 2005.

[84] D. Deutsch. Quantum theory, the Church-Turing principle and the universal quantum computer. *Proceedings of the Royal Society of London; Series A, Mathematical and Physical Sciences*, 400(1818):97–117, 1985.

[85] P. Diaconis. Patterns in eigenvalues: The 70th Josiah Willard Gibbs lecture. *Bulletin of the American Mathematical Society*, 40:155–178, 2003.

[86] D. Dieks. Communication by EPR devices. *Physical Letters A*, 92(6):271–272, 1982.

[87] D. DiVincenzo, M. Horodecki, D. W. Leung, J. A. Smolin, and B. M. Terhal. Locking classical correlations in quantum states. *Physical Review Letters*, 92(6):067902, Feb 2004.

[88] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri. Complete family of separability criteria. *Physical Review A*, 69:022308, 2004.

[89] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47(10):777–780, 1935.

[90] M. Fannes and C. Vandenplas. Finite size mean-field models. *Journal of Physics A - Mathematical and General*, 39:13843–13860, 2006.

[91] U. Fano. Pairs of two-level systems. *Reviews in Modern Physics*, 55:855–874, 1983.

[92] E. Farhi, J. Goldstone, and S. Gutmann. A quantum algorithm for the Hamiltonian NAND tree. *Theory of Computing*, 4(8):169–190, 2008.

[93] B. Fefferman and C. Umans. Pseudorandom generators and the BQP vs. PH problem. Available at arXiv.org e-Print quant-ph/1007.0305v3, 2010.

[94] A. Ferraro, L. Aolita, D. Cavalcanti, F. M. Cucchietti, and A. Acín. Almost all quantum states have nonclassical correlations. *Physical Review A*, 81:052318, 2010.

[95] R. Feynman. *The Feynman Lectures on Physics*, volume III. Addison-Wesley, 1964.

[96] R. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6–7):467–488, 1982.

[97] R. Feynman. Quantum mechanical computers. *Optics News*, 11:11, 1985.

[98] L. Fortnow. Time-space tradeoffs for satisfiability. *Journal of Computer and System Sciences*, 60(2):337–353, 2000.

[99] L. Fortnow, R. Lipton, D. van Melkebeek, and A. Viglas. Time-space lower bounds for satisfiability. *Journal of the ACM*, 52:835–865, 2005.

[100] L. Fortnow and J. Rogers. Complexity limitations on quantum computation. *Journal of Computer and System Sciences*, 59(2):240–252, 1999.

[101] A. M. Frieze and R. Kannan. The regularity lemma and approximation schemes for dense problems. In *Proceedings of the 37th Annual IEEE Symposium on Foundations of Computer Science (FOCS 1996)*, pages 12–20, 1996.

[102] L. Fu. Nonlocal effect of a bipartite system induced by local cyclic operation. *Europhys. Lett.*, 75:1, 2006.

[103] F. Le Gall, S. Nakagawa, and H. Nishimura. On QMA protocols with two short quantum proofs. *Quantum Information & Computation*, 12(7&8):0589–0600, 2012.

[104] S. Gharibian. QMA-completeness of the 5-local Hamiltonian problem, 2009. Course project for CS898: Quantum complexity theory. Course webpage: http://www.cs.uwaterloo.ca/~watrous/qcomplexity/.

[105] S. Gharibian. Strong NP-hardness of the quantum separability problem. *Quantum Information and Computation*, 10(3&4):343–360, 2010.

[106] S. Gharibian. Quantifying non-classicality with local unitary operations. Available at arXiv.org e-Print quant-ph/1202.1598v1, 2012.

[107] S. Gharibian, H. Kampermann, and D. Bruß. On global effects caused by locally noneffective unitary operations. *Quantum Information & Computation*, 9:1013–1029, 2008.

[108] S. Gharibian and J. Kempe. Approximation algorithms for QMA-complete problems. In *Proceedings of 26th IEEE Conference on Computational Complexity (CCC 2011)*, pages 178–188, 2011. DOI: 10.1109/CCC.2011.15, © 2011 IEEE, ieeexplore.ieee.org.

[109] S. Gharibian and J. Kempe. Hardness of approximation for quantum problems. In *Proceedings of 39th International Colloquium on Automata, Languages and Programming (ICALP 2012)*, pages 387–398, 2012. DOI: 10.1007/978-3-642-31594-7, © 2012 Springer, www.springerlink.com.

[110] S. Gharibian, M. Piani, G. Adesso, J. Calsamiglia, and P. Horodecki. Characterizing quantumness via entanglement creation. *International Journal of Quantum Information*, 9(7 & 8):1701–1713, 2011. DOI: 10.1142/S0219749911008258, © 2011 World Scientific Publishing Company, www.worldscientific.com/worldscinet/ijqi.

[111] S. Gharibian, J. Sikora, and S. Upadhyay. QMA variants with polynomially many provers. Available at arXiv.org e-Print quant-ph/1108.0617v1, 2011.

[112] S. M. Giampaolo and F. Illuminati. Characterization of separability and entanglement in (2D)- and (3D)-dimensional systems by single-qubit and single-qutrit unitary transformations. *Physical Review A*, 76(4):042301, 2007.

[113] D. Gillman. A Chernoff bound for random walks on expanders. In *Proceedings of the 34th Annual IEEE Symposium on the Foundations of Computer Science (FOCS 1993)*, pages 680–691, 1993.

[114] D. Girolami and G. Adesso. Interplay between computable measures of entanglement and other quantum correlations. *Physical Review A*, 84:052110, 2011.

[115] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45(4):653–750, 1998.

[116] O. Goldreich and D. Zuckerman. Another proof that BPP $\subseteq$ PH (and more). *Electronic Colloquium on Computational Complexity*, 1997.

[117] D. Gottesman and S. Irani. The quantum and classical complexity of translationally invariant tiling and Hamiltonian problems. In *Proceedings of the 50th IEEE Symposium on Foundations of Computer Science*, pages 95–104, 2009. Full version available at arXiv.org e-Print quant-ph/0905.2419v2.

[118] B. Groisman, D. Kenigsberg, and T. Mor. "Quantumness" versus "classicality" of quantum states. Available at arXiv.org e-Print quant-ph/0703103, 2007.

[119] B. Groisman, S. Popescu, and A. Winter. Quantum, classical, and total amount of correlations in a quantum state. *Physical Review A*, 72(3):032317, 2005.

[120] D. Gross, S. Flammia, and J. Eisert. Most quantum states are too entangled to be useful as computational resources. *Physical Review Letters*, 102:190501, 2009.

[121] M. Grötschel, L. Lovàsz, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer-Verlag, 1993.

[122] L. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the 28th ACM Symposium on the Theory of Computing (STOC 1996)*, pages 212–219, 1996.

[123] L. Gurvits. Classical deterministic complexity of Edmond's problem and quantum entanglement. In *Proceedings of the 35th Symposium on Theory of computing*, pages 10–19. ACM Press, 2003.

[124] L. Gurvits and H. Barnum. Largest separable balls around the maximally mixed bipartite quantum state. *Physical Review A*, 66(6):062311, 2002.

[125] L. Gurvits and H. Barnum. Separable balls around the maximally mixed multipartite quantum states. *Physical Review A*, 68(4):042312, 2003.

[126] L. Gurvits and H. Barnum. Better bound on the exponent of the radius of the multipartite separable ball. *Physical Review A*, 72(3):032322, 2005.

[127] G. Gutoski. *Quantum strategies and local operations*. PhD Thesis, University of Waterloo, 2009. Available at arXiv.org e-Print quant-ph/1003.0038.

[128] A. Harrow and A. Montanaro. An efficient test for product states, with applications to quantum Merlin-Arthur games. In *Proceedings of the 51st IEEE Annual Symposium on Foundations of Computer Science*, pages 633–642, 2010.

[129] A. W. Harrow, A. Hassadim, and S. Lloyd. Quantum algorithm for solving linear systems of equations. *Physical Review Letters*, 15(103):150502, 2009.

[130] G. Hast. Approximating Max kCSP - outperforming a random assignment with almost a linear factor. In *Proceedings of the 32nd International Colloquium on Automata, Languages, and Programming (ICALP 2005)*, pages 956–968, 2005.

[131] J. Håstad. Some optimal inapproximability results. In *Proceedings of the 29th Symposium on Theory of Computing (STOC 1997)*, pages 1–10, 1997.

[132] J. Håstad. On the approximation resistance of a random predicate. In *Lecture Notes in Computer Science*, volume 4627, pages 149–163, 2007.

[133] M. Hastings. An area law for one-dimensional quantum systems. *Journal of Statistical Mechanics*, P08024(08), 2007.

[134] M. B. Hastings. Trivial low energy states for commuting hamiltonians, and the quantum PCP conjecture. Available at arXiv.org e-Print quant-ph/1201.3387, 2012.

[135] P. Hausladen and W. K. Wootters. A 'pretty good measurement' for distinguishing quantum states. *Journal of Modern Optics*, 41(12):2385–2390, 1994.

[136] P. Hayden, D. Leung, and A. Winter. Aspects of generic entanglement. *Communications in Mathematical Physics*, 265(1):95–117, 2006.

[137] P. Hayden, D. W. Leung, P. Shor, and A. Winter. Randomizing quantum states: Constructions and applications. *Communications in Mathematical Physics*, 250(371), 2004.

[138] L. Henderson and V. Vedral. Classical, quantum and total correlations. *Journal of Physics A - Mathematical and General*, 34:6899, 2001.

[139] F. T. Hioe and J. H. Eberly. N-level coherence vector and higher conservation laws in quantum optics and quantum mechanics. *Physical Review Letters*, 47:838, 1981.

[140] T. Hiroshima and M. Hayashi. Finding a maximally correlated state: Simultaneous schmidt decomposition of bipartite pure states. *Physical Review A*, 70(3):030302, 2004.

[141] D. Hochbaum. *Approximation Algorithms for NP-Hard Problems*. Wadsworth Publishing Company, 1997.

[142] W. Höffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1964.

[143] R. A. Horn and C. H. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

[144] M. Horodecki, P. Horodecki, and R. Horodecki. Mixed-state entanglement and distillation: is there a "bound" entanglement in nature? *Physical Review Letters*, 80:5239–5242, 1998.

[145] M. Horodecki, P. Horodecki, R. Horodecki, J. Oppenheim, A. Sen De, U. Sen, and B. Synak. Local versus non-local information in quantum information theory: formalism and phenomena. *Physical Review A*, 71:062307, 2005.

[146] Michal Horodecki, Pawel Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physical Letters A*, 223(1–2):1–8, 1996.

[147] P. Horodecki. Separability criterion and inseparable mixed states with positive partial transposition. *Physical Letters A*, 232:333, 1997.

[148] R. Horodecki and M. Horodecki. Information-theoretic aspects of quantum inseparability of mixed states. *Physical Review A*, 54(3):1838–1843, 1996.

[149] R. Horodecki and P. Horodecki. Perfect correlations in the Einstein-Podolsky-Rosen experiment and Bell's inequalities. *Physical Letters A*, 210:227, 1996.

[150] R. Horodecki, P. Horodecki, and M. Horodecki. Violating Bell inequality by mixed spin-1/2 states: necessary and sufficient condition. *Physical Letters A*, 200:340–344, 1995.

[151] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Reviews of Modern Physics*, 81(2):865–942, 2009.

[152] M. Żukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert. "Event-ready-detectors" bell experiment via entanglement swapping. *Physical Review Letters*, 71(26):4287–4290, Dec 1993.

[153] Karol Życzkowski, Paweł Horodecki, Anna Sanpera, and Maciej Lewenstein. Volume of the set of separable states. *Physical Review A*, 58(2):883–892, Aug 1998.

[154] L. Ioannou. Computational complexity of the quantum separability problem. *Quantum Information & Computation*, 7(4):335, 2007.

[155] R. Jain, Z. Ji, S. Upadhyay, and J. Watrous. QIP = PSPACE. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*, pages 573–581, 2010.

[156] D. Janzing and P. Wocjan. BQP-complete problems concerning mixing properties of classical random walks on sparse graphs. Available at arXiv.org e-Print quant-ph/0610235v2, 2006.

[157] D. Janzing, P. Wocjan, and T. Beth. "Non-Identity-Check" is QMA-complete. *International Journal of Quantum Information*, 3:463–473, 2005.

[158] S. P. Jordan, D. Gosset, and P. J. Love. Quantum-Merlin-Arthur-complete problems for stoquastic Hamiltonians and Markov matrices. *Physical Review A*, 81:032331, 2010.

[159] S. P. Jordan, H. Kobayashi, D. Nagaj, and H. Nishimura. Achieving perfect completeness in classical-witness quantum Merlin-Arthur proof systems. *Quantum Information & Computation*, 12(5 & 6):461–471, 2012.

[160] R. Jozsa and N. Linden. On the role of entanglement in quantum-computational speed-up. *Proceedings of the Royal Society of London; Series A, Mathematical and Physical Sciences*, 459:2011–2032, 2003.

[161] A. Kay. Quantum-Merlin-Arthur-complete translationally invariant Hamiltonian problem and the complexity of finding ground-state energies in physical systems. *Physical Review A*, 76(3):030307, 2007.

[162] P. Kaye, R. Laflamme, and M. Mosca. *An Introduction to Quantum Computing*. Oxford University Press, 2007.

[163] J. Kempe, A. Kitaev, and O. Regev. The complexity of the local Hamiltonian problem. *SIAM Journal on Computing*, 35(5):1070–1097, 2006.

[164] J. Kempe and O. Regev. 3-local Hamiltonian is QMA-complete. *Quantum Information & Computation*, 3(3):258–264, 2003.

[165] J. Kempe, O. Regev, and B. Toner. Unique games with entangled provers are easy. *SIAM Journal on Computing*, 39(7):3207–3229, 2010.

[166] D. Kenigsberg, T. Mor, and G. Ratsaby. Quantum advantage without entanglement. *Quantum Information & Computation*, 6:606, 2006.

[167] S. Khanna, M. Sudan, L. Trevisan, and D. Williamson. The approximability of constraint satisfaction problems. *SIAM Journal on Computing*, 30(6):1863–1920, 2001.

[168] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th Symposium on Theory of Computing (STOC 2002)*, pages 767–775, 2002.

[169] G. Kimura. The Bloch vector for N-level systems. *Physical Letters A*, 314(5), August 2003.

[170] A. Kitaev. Quantum NP, 1999. Talk at Second Workshop on Algorithms in Quantum Information Processing (AQIP 1999), DePaul University.

[171] A. Kitaev, A. Shen, and M. Vyalyi. *Classical and Quantum Computation*. American Mathematical Society, 2002.

[172] A. Kitaev and J. Watrous. Parallelization, amplification, and exponential time simulation of quantum interactive proof systems. In *Proceedings of the 32nd ACM Symposium on Theory of Computing (STOC 2000)*, pages 608–617, 2000.

[173] A. Y. Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191–1249, 1997.

[174] E. Knill and R. Laflamme. Power of one bit of quantum information. *Physical Review Letters*, 81:5672, 1998.

[175] H. Kobayashi, K. Matsumoto, and T. Yamakami. Quantum Merlin Arthur proof systems: Are multiple Merlins more helpful to Arthur? In *Proceedings of the 14th International Symposium on Algorithms and Computation*, pages 189–198, 2003. Volume 2906 of *Lecture Notes in Computer Science*, Springer.

[176] C. R. Laumann, A. M. Läuchli, R. Moessner, A. Scardicchio, and S. L. Sondhi. Product, generic, and random generic quantum satisfiability. *Physical Review A*, 81:062345, 2010.

[177] C. Lautemann. BPP and the polynomial time hierarchy. *Information Processing Letters*, 17:215–218, 1983.

[178] T. Lee, R. Mittal, B. W. Reichardt, R. Spalek, and M. Szegedy. Quantum query complexity of state conversion. In *Proceedings of the 52nd IEEE Symposium on Foundations of Computer Science (FOCS 2011)*, pages 344–353, 2011.

[179] L. Levin. Universal search problems. *Problems of Information Transmission*, 9(3):265–266, 1973.

[180] M. Lewin, D. Livnat, and U. Zwick. Improved rounding techniques for MAX 2-SAT and MAX DI-CUT problems. In *Proceedings of the 9th International IPCO Conference on Integer Programming and Combinatorial Optimization (IPCO 2002)*, pages 67–82, 2002.

[181] B. Li, S.-M. Fei, Z.-X. Wang, and H. Fan. Assisted state discrimination without entanglement. *Physical Review A*, 85:022328, 2012.

[182] Y.-K. Liu. Consistency of local density matrices is QMA-complete. In *Lecture Notes in Computer Science*, volume 4110, pages 438–449, 2006.

[183] Y.-K. Liu. The local consistency problem for stoquastic and 1-D quantum systems. Available at arXiv.org e-Print quant-ph/0712.1388v2, 2007.

[184] Y.-K. Liu, M. Christandl, and F. Verstraete. Quantum computational complexity of the N-representability problem: QMA complete. *Physical Review Letters*, 98:110503, 2007.

[185] S. Luo. Using measurement-induced disturbance to characterize correlations as classical or quantum. *Physical Review A*, 77:022301, 2008.

[186] S. Luo and S. Fu. Geometric measure of quantum discord. *Physical Review A*, 82:034302, 2010.

[187] M. Horodecki and P. Horodeki and R. Horodecki and J. Oppenheim and A. Sen De and U. Sen and B. Synak. Local versus non-local information in quantum-information theory: Formalism and phenomena. *Physical Review A*, 71(6):062307, 2005.

[188] M. Piani and M. Christandl and C. E. Mora and P. Horodecki. Broadcast copies reveal the quantumness of correlations. *Physical Review Letters*, 102(25):250503, 2009.

[189] Hans Maassen and J. B. M. Uffink. Generalized entropic uncertainty relations. *Physical Review Letters*, 60(12):1103–1106, 1988.

[190] V. Madhok and A. Datta. Interpreting quantum discord through quantum state merging. *Physical Review A*, 83:032323, 2011.

[191] C. Marriott and J. Watrous. Quantum Arthur-Merlin games. *Computational Complexity*, 14(2):122–152, 2005.

[192] L. Masanes. All bipartite entangled states are useful for information processing. *Physical Review Letters*, 96:150501, 2006.

[193] A. Meyer and L. Stockmeyer. The equivalence problem for regular expressions with squaring requires exponential time. In *Proceedings of the 13th Symposium on Foundations of Computer Science*, pages 125–129, 1972.

[194] D. A. Meyer. Sophisticated quantum search without entanglement. *Physical Review Letters*, 85:2014, 2000.

[195] K. Modi, A. Brodutch, H. Cable, T. Paterek, and V. Vedral. Quantum discord and other measures of quantum correlation. Available at arXiv.org e-Print quant-ph/1112.6238v1, 2011.

[196] K. Modi, T. Paterek, W. Son, V. Vedral, and M. Williamson. Unified view of quantum and classical correlations. *Physical Review Letters*, 104:080501, 2010.

[197] A. Monras, G. Adesso, S. M. Giampaolo, G. Gualdi, G. B. Davies, and F. Illuminati. Entanglement quantification by local unitaries. *Physical Review A*, 84:012301, 2011.

[198] D. Nagaj. *Local Hamiltonians in Quantum Computation*. PhD thesis, Massachusetts Institute of Technology, Boston, 2008. Available at arXiv.org e-Print quant-ph/0808.2117v1.

[199] D. Nagaj and S. Mozes. A new construction for a QMA complete 3-local Hamiltonian. *Journal of Mathematical Physics*, 48(7):072104, 2007.

[200] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[201] M. A. Nielsen and J. Kempe. Separable states are more disordered globally than locally. *Physical Review Letters*, 86:5184–5187, 2001.

[202] R. Oliveira and B. M. Terhal. The complexity of quantum spin systems on a two-dimensional square lattice. *Quantum Information & Computation*, 8(10):0900–0924, 2008.

[203] H. Ollivier and W. H. Zurek. Quantum discord: A measure of the quantumness of correlations. *Physical Review Letters*, 88:017901, 2002.

[204] T. J. Osborne. Hamiltonian complexity. Available at arXiv.org e-Print quant-ph/1106.5875v1, 2011.

[205] S. Östlund and S. Rommer. Thermodynamic limit of density matrix renormalization. *Physical Review Letters*, 75:3537–3540, 1995.

[206] Asher Peres. Separability criterion for density matrices. *Physical Review Letters*, 77(8):1413–1415, Aug 1996.

[207] I. Peschel, X. Wang, M. Kaulke, and K. Hallberg (Edgs.). Density-matrix renormalization - a new numerical method in physics. In *Lecture Notes in Physics*, volume 528. Springer-Verlag, 1998.

[208] M. Piani, S. Gharibian, G. Adesso, J. Calsamiglia, P. Horodecki, and A. Winter. All non-classical correlations can be activated into distillable entanglement. *Physical Review Letters*, 106:220403, 2011. DOI: 10.1103/PhysRevLett.106.220403, © 2011 American Physical Society, prl.aps.org.

[209] Marco Piani, Paweł Horodecki, and Ryszard Horodecki. No-local-broadcasting theorem for multipartite quantum correlations. *Physical Review Letters*, 100(9):090502, 2008.

[210] Martin B. Plenio and S. Virmani. An introduction to entanglement measures. *Quantum Information & Computation*, 7:1–51, 2007.

[211] P. Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the 40th ACM Syposium on Theory of Computing (STOC 2008)*, pages 245–254, 2008.

[212] E. Rains. A semidefinite program for distillable entanglement. *IEEE Transactions on Information Theory*, 47(7):2921–2933, 2001.

[213] L. Roa, J. C. Retamal, and M. Alid-Vaccarezza. Dissonance is required for assisted optimal state discrimination. *Physical Review Letters*, 107:080401, 2011.

[214] S. Rommer and S. Östlund. Class of ansatz wave functions for one-dimensional spin systems and their relation to the density matrix renormalization group. *Physical Review B*, 55:2164–2181, 1997.

[215] B. Rosgen. Testing non-isometry is QMA-complete. In *Proceedings of the 5th conference on Theory of quantum computation, communication, and cryptography (TQC 2010)*, pages 63–76, 2010.

[216] R. Rossignoli, N. Canosa, and L. Ciliberti. Generalized entropic measures of quantum correlations. *Physical Review A*, 82(5):052342, 2010.

[217] Akira SaiToh, Robabeh Rahimi, and Mikio Nakahara. Nonclassical correlation in a multipartite quantum system: Two measures and evaluation. *Physical Review A*, 77(5):052101, 2008.

[218] A. Samorodnitsky and L. Trevisan. Gowers uniformity, influences of variables, and PCPs. In *Proceedings of the 38th Symposium on Theory of Computing (STOC 2006)*, pages 11–20, 2006.

[219] M. Schaefer and C. Umans. SIGACT news complexity theory column 38. In L. Hemaspaandra, editor, *ACM SIGACT News*, volume 33. 2002.

[220] U. Schollwöck. The density-matrix renormalization group. *Reviews in Modern Physics*, 77:259–315, 2005.

[221] E. Schrödinger. Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften*, 23(48):807–812, 1935.

[222] N. Schuch and J. I. Cirac. Matrix product state and mean-field solutions for one-dimensional systems can be found efficiently. *Physical Review A*, 82:012314, 2010.

[223] N. Schuch and F. Verstraete. Computational complexity of interacting electrons and fundamental limitations of density functional theory. *Nature Physics*, 5:732–735, 2009.

[224] P. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997.

[225] D. R. Simon. On the power of quantum computation. In *Proceedings of the 35th Symposium on Foundations of Computer Science (FOCS 1994)*, pages 116–123, 1994.

[226] D. R. Simon. On the power of quantum computation. *SIAM Journal on Computing*, 26(5):1474–1483, 1997.

[227] M. Sipser. A complexity theoretic approach to randomness. In *Proceedings of the 15th Symposium on Theory of computing*, pages 330–335. ACM Press, 1983.

[228] A. Srinivasan and D. Zuckerman. Computing with very weak random sources. In *Proceedings of the 35th Symposium on Foundations of Computer Science*, pages 264–275, 1994.

[229] A. Streltsov, S. M. Giampaolo, W. Roga, D. Bruß, and F. Illuminati. Nonlocality of quantum correlations. Available at arXiv.org e-Print quant-ph/1206.4075v2, 2012.

[230] A. Streltsov, H. Kampermann, and D. Bruß. Quantum cost for sending entanglement. Available at arXiv.org e-Print quant-ph/1203.1264v2, 2012.

[231] Alexander Streltsov, Hermann Kampermann, and Dagmar Bruß. Linking quantum discord to entanglement in a measurement. *Physical Review Letters*, 106:160401, 2011.

[232] A. Ta-Shma, C. Umans, and D. Zuckerman. Lossless condensers, unbalanced expanders, and extractors. *Combinatorica*, 27(2):213–240, 2007.

[233] S. Tamaryan, T.-C. Wei, and D. Park. Maximally entangled three-qubit states via geometric measure of entanglement. *Physical Review A*, 80:052315, 2009.

[234] L. Trevisan. Parallel approximation algorithms by positive linear programming. *Algorithmica*, 21(1):72–88, 1998.

[235] C. Umans. Hardness of approximating Σ_2^p minimization problems. In *Proceedings of the 40th Symposium on Foundations of Computer Science*, pages 465–474, 1999.

[236] V. Vazirani. *Approximation Algorithms*. Springer, 2001.

[237] V. Vedral. The role of relative entropy in quantum information theory. *Reviews in Modern Physics*, 74:197–234, 2002.

[238] V. Vedral and M. B. Plenio. Entanglement measures and purification procedures. *Physical Review A*, 57(3):1619–1633, 1998.

[239] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight. Quantifying entanglement. *Physical Review Letters*, 78(12):2275–2279, Mar 1997.

[240] G. Vidal. Efficient classical simulation of slightly entangled quantum computations. *Physical Review Letters*, 91:147902, 2003.

[241] G. Vidal and R. F. Werner. A computable measure of entanglement. *Physical Review A*, 65:032314, 2002.

[242] S. Vinjanampathy and A. R. P. Rau. Calculation of quantum discord for qubit-qudit or N qubits. *Journal of Physics A: Mathematical and Theoretical*, 45:095303, 2012.

[243] J. Watrous. Private communication.

[244] J. Watrous. Limits on the power of quantum statistical zero-knowledge. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 459–468, 2002.

[245] J. Watrous. Lecture 1: Mathematical Preliminaries Part I, 2008. Latest version available at: www.cs.uwaterloo.ca/~watrous/CS766/.

[246] J. Watrous. Lecture 2: Mathematical Preliminaries Part II, 2008. Latest version available at: www.cs.uwaterloo.ca/~watrous/CS766/.

[247] J. Watrous. Lecture 5: Naimark's Theorem; Characterization of quantum operations, 2008. Latest version available at: www.cs.uwaterloo.ca/~watrous/CS766/.

[248] J. Watrous. *Encyclopedia of Complexity and System Science*, chapter Quantum Computational Complexity. Springer, 2009.

[249] J. Watrous. Semidefinite programs for completely bounded norms. *Theory of Computing*, 5:217–238, 2009.

[250] J. Watrous. Lecture 14: Separable operators, 2011. Latest version available at: www.cs.uwaterloo.ca/~watrous/CS766/.

[251] S. Wehner and A. Winter. Entropic uncertainty relations - a survey. *New Journal of Physics - Special Issue on Quantum Information and Many-Body Theory*, 12:025009, 2010.

[252] T.-C. Wei and P. M. Goldbart. Geometric measure of entanglement and applications to bipartite and multipartite quantum states. *Physical Review A*, 68:042307, 2003.

[253] T.-C. Wei, M. Mosca, and A. Nayak. Interacting boson problems are QMA-hard. *Physical Review Letters*, 104:040501, 2010.

[254] R. F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Physical Review A*, 40(8):4277–4281, 1989.

[255] S. R. White. Density matrix formulation for quantum renormalization groups. *Physical Review Letters*, 69:2863–2866, 1992.

[256] S. R. White. Density-matrix algorithms for quantum renormalization groups. *Physical Review B*, 48:10345–10356, 1993.

[257] P. Wocjan and J. Yard. The Jones polynomial: Quantum algorithms and applications in quantum complexity theory. *Quantum Information and Computation*, 8(1&2):0147–0180, 2008.

[258] W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.

[259] S. Wu, E. V. Poulsen, and K. Mølmer. Correlations in local measurements on a quantum state, and complementarity as an explanation of nonclassicality. *Physical Review A*, 80:032319, 2009.

[260] T. Yamakami. Quantum NP and a quantum hierarchy. In *Proceedings of the 2nd IFIP International Conference on Theoretical Computer Science*, pages 323–336. Kluwer Academic Publishers, 2002.

[261] H. P. Yuen. Amplification of quantum states and noiseless photon amplifiers. *Physical Letters A*, 113:405–407, 1986.

[262] S. Zachos and M. Furer. Probabalistic quantifiers vs. distrustful adversaries. In *Foundations of Software Technology and Theoretical Computer Science, 7th Conference*, pages 443–455, 1987. Volume 287 of *Lecture Notes in Computer Science*.

[263] D. Zuckerman. On unapproximable versions of NP-complete problems. *SIAM Journal on Computing*, 25(6):1293–1304, 1996.